THE EFFECT OF HELICASES ON THE INSTABILITY OF CTG·CAG TRINUCLEOTIDE REPEAT ARRAYS IN THE *ESCHERICHIA COLI* CHROMOSOME

ADAM JACKSON

Thesis presented for the degree of Doctor of Philosophy

The University of Edinburgh

June 2010

Declaration

I hereby declare that this thesis was composed by me and the work described is my own, unless otherwise stated. This work has not been submitted for any other degree or professional qualification. Information obtained from sources other than this study is acknowledged in the text or included in the references.

Adam Jackson

June 2010

“If you’re not part of the solution, you’re part of the precipitate.”

Anon

Acknowledgements

I would first of all like to thank David Leach for allowing me to work in his lab, and providing so much help and guidance. My thanks also go to the Medical Research Council for funding my PhD. I would like to thank Jill Sales and Suzy Moat for help and discussions about statistics. Elise Darmon for critical reading of this thesis, and Martin White and Federica Andreoni for discussions about the work presented herein. In fact, I would like to thank all the members of the leach lab, past and present and the whole forth floor, for the many coffee breaks and cakes. The support and friendship of the people I have worked with over the last four years has been amazing, whether it be discussing work when encouragement and advice was needed, or talking non-sense when I really just needed to smile. Thank you all for the help you have given me.

I would like to thank everyone in the Genepool for all their help with this thesis and putting up with me for the last few years sending so many GeneMapper plates to you!

I would also like to thank all my friends in Liverpool who helped me get to where I am today, helped me make some difficult decisions about the direction of my life and encouraged me to find my own way. I wish to thank TESS for all the help in forgetting the stresses of the day, providing a drum to hit, a beat to lift me, and normally a pint or two to settle me back down.

Most of all this thesis is dedicated to my Mum and Dad, who have always been my biggest support, always believing in me, even when I didn't believe in myself.

Abstract

A trinucleotide repeat (TNR) is a 3 base pair (bp) DNA sequence tandemly repeated in an array. In humans, TNR sequences have been found to be associated with at least 14 severe neurological diseases including Huntington disease, myotonic dystrophy and several of the spinocerebellar ataxias. Such diseases are caused by an expansion of the repeat sequence beyond a threshold length and are characterized by non-Mendelian patterns of inheritance which lead to genetic anticipation. Although the mechanism of the genetic instability in these arrays is not yet fully understood, various models have been suggested based on the *in vitro* observation that TNR sequences can form secondary structures such as pseudo-hairpins.

In order to investigate the mechanisms responsible for instability of TNR sequences, a study was carried out on *Escherichia coli* cells in which TNR arrays had been integrated into the chromosomal *lacZ* gene. This genetic assay was used to identify proteins and pathways involved in deletion and/or expansion instability.

Deletion instability was clearly dependent on orientation of the TNR sequence relative to the origin of replication. Interestingly, it was found that expansion instability is not dependent on the orientation of the repeat array relative to the origin of replication.

The replication fork reversal pathway and the RecFOR mediated gap repair pathway were found to have no statistically significant influence on the

instability of TNR arrays. However, the protein UvrD was found to affect the deletion instability of TNR sequences.

The roles of key helicase genes were investigated for their effects on instability of chromosomal CTG•CAG repeats. Mutation of the *rep* gene increased deletion in the CTG leading-strand orientation of the repeat array, and expansion in both orientations - destabilizing the TNR array. RecQ helicase was found to have a significant effect on TNR instability in the orientation in which CAG repeats were present on the leading-strand relative to the origin of replication. Mutation of the *recQ* gene severely limited the number of expansion events in this orientation, whilst having no effect on deletions. This dependence of expansions on RecQ was lost in a *rep* mutant strain. In a *rep* mutant expansions were shown to be partially dependent on the DinG helicase.

All together, these results suggest a model of TNR instability in which expansions are due to events occurring at either the leading or lagging strand of an arrested replication fork, facilitated by helicase action. The identity of the helicase implicated is determined by the nature of the arrest.

Abbreviations

Abbreviation	
ATP	Adenosine Triphosphate
Bp	Base Pair
DM	Myotonic Dystrophy
DNA	Deoxyribonucleic Acid
dNTP	Deoxyribonucleoside Triphosphate
dsDNA	Double Stranded DNA
G	Gram
HJ	Holliday Junction
HR	Homologous Recombination
Kb	Kilo base
L	Litre
LB	Luria Broth
M	Molar
NHEJ	Non-Homologous End Joining
PCR	Polymerase Chain Reaction
PM	Pre-Mutation
RFR	Replication Fork Reversal
RNA	Ribonucleic Acid
ssDNA	Single Stranded DNA
T_m	Melting Temperature
TNR	Trinucleotide Repeat
TRED	Trinucleotide Repeat Expansion Disease
UTR	Untranslated Region
YAC	Yeast Artificial Chromosome

Table Of Contents

Chapter Two: Materials and Methods

Chapter Six: Concluding Remarks

Table of Tables

Table of Figures

Chapter One:

Introduction

1.1 Introduction

Deoxyribonucleic acid (DNA) holds all the genetic information of a cell. Containing the four bases adenine, guanine, cytosine and thymine, the DNA stores the blueprints for, and regulates the production of, all the proteins produced by an organism. As an organism grows, and passes its genetic information from one cell to another, the faithful transmission of that exact code, maintaining the stability of the genome, is required for the new cell to function as the previous cell. A change to the DNA sequence can in some cases, be beneficial to an organism, as the molecular basis for evolution, however it may also be unfavourable – harming the cell and the organism.

Throughout the genome of every organism sequenced to date are examples of regions of DNA in which the distribution of bases is in repetitive tracts. The smallest of these tracts have been termed microsatellites, they contain a repeat unit of 1-6bp in length, and are found in both prokaryotic and eukaryotic organisms. Instability of the genome arising in such repetitive sequences has been found to be responsible for many diseases in human beings (Lopez Castel *et al.*, 2010; Mirkin, 2007; Pearson *et al.*, 2005). One group of microsatellite sequences that have attracted much attention due to their role in human diseases are trinucleotide repeat (TNR) sequences. TNRs consist of a repeat unit of 3bp, repeated in tandem. Three of the TNR sequences have been associated with disease in humans, these are CAG, CTG and GAA. Instability of TNR sequences has been described as 'dynamic mutation', as the mutability of the sequence is linked to the copy number of the repeat unit. Therefore the

product of a mutation event, altering the number of repeat units, will have a different potential mutability than its parent (Richards and Sutherland, 1992).

1.2 Repeat sequences and disease

1.2.1 Trinucleotide Repeat Expansion Diseases (TRED)

Expanded TNR sequences have been shown to be the cause of at least 14 severe neurological disorders in human beings (Orr and Zoghbi, 2007), termed TREDs – Trinucleotide Repeat Expansion Disorders. They show non-Mendalian pattern of disease due to the common feature of genetic anticipation of the TNR sequence (Pearson *et al.*, 2005), in which the disease shows a younger age of onset as it is inherited.

Disease causing TNR arrays have been found in a variety of different loci in the human genome, in intronic regions of genes, in the 5'- and 3'-Untranslated regions (UTR), or in coding regions of genes – allowing sub-categorization of the diseases depending on whether the TNR sequence is in a coding region or not.

1.2.2 TNR in non-coding regions

Non-coding TREDs have been found with the TNR sequence located in the 5'-UTR (fragile X syndrome), the 3'-UTR (myotonic dystrophy) and in intronic

regions within a gene (Friedreich Ataxia). Disease causing TNR sequences in these loci tend to be larger than those in the coding regions of a gene (Orr and Zoghbi, 2007).

1.2.2.1 Fragile X Syndrome

The first link between disease and TNR sequences was made in 1991 by Verkerk and colleagues when investigating the disease Fragile X syndrome (Verkerk *et al.*, 1991). Fragile X syndrome is the major inherited cause of mental retardation in males. This disease is an X linked dominant disorder characterized by mental retardation – with the average IQ of sufferers being 20-60, as well as other symptoms such as hyperactivity, autism-like behaviours and attention deficit disorder. Anatomical characteristics of the disease include the presence of a prominent jaw, large ears and enlarged testicles (Kaytor and Orr, 2001). As the name would suggest individuals with the syndrome display a fragile site in the X chromosome mapped to Xq27.3 (Krawczun *et al.*, 1985).

Verkerk found that an expanded CGG repeat array in the 5'-UTR of the gene *FMR1* was associated with the disease state (Verkerk *et al.*, 1991). It has since been shown that this CGG TNR array is accompanied by hypermethylation of the CpG island it presents, followed by histone deacetylation, which cause transcriptional silencing of the gene (D'Hulst and Kooy, 2009). The *FMR*1 gene encodes the FMRP (Fragile X Mental Retardation Protein) RNA binding protein, which normally works in brain

neuronal cells to regulate translation within the dendrites. The absence of this protein due to silencing of the gene is the cause of the symptoms of the disease state (Brouwer *et al.*, 2009).

TREDs are caused by a TNR array expanded beyond a threshold length. TNR sequences below the threshold length, which can vary for each disease locus, are relatively stable. However, once the threshold length has been achieved the sequence becomes much less stable and the probability of expansion instability occurring increases (Kovtun and McMurray, 2008). The CGG repeat associated with the *FMR*1 gene in disease free individuals is around 29-30 repeats long, while the threshold length for the disease, at which transcription can be silenced, is 200 repeat units. Fragile X syndrome and a number of other TREDs also display an intermediate genotype, in this case between 50-200 repeat units, at which carriers do not display the phenotype of the disease, rather they are said to be pre-mutation (PM) carriers of the disease (Van Esch, 2006).

PM carriers have been shown however to develop other conditions believed to be due to the expanded TNR sequences they carry. Fragile X-associated tremor/ataxia syndrome is found mainly in older male carriers and is characterized by the presence of ubiquitin-positive intranuclear inclusions in brain neurons and astrocytes (Greco *et al.*, 2002). PM carriers of *FMR1* gene have been shown to have elevated levels of *FMR1* mRNA but normal levels of the FMRP protein. A proposed molecular model of the disease involves a gain

of function mutation of the mRNA, similar to myotonic dystrophy (reviewed in (Brouwer *et al.*, 2009)).

1.2.2.2 Myotonic Dystrophy

Myotonic Dystrophy is a multi-systemic disorder that affects 1:8000 people worldwide. The phenotypic signs of the disease include muscle weakness, wasting and myotonia (a difficulty in relaxation of muscle). There are two forms of the disease, DM1 and DM2. DM1 is caused by an expanded CTG repeat array in the 3'-UTR of the gene *DMPK*, which encodes a protein kinase. The DM2 form of the disease is due to an expanded tetranucleotide repeat array (CCTG) in an intron of the gene *ZNF*9, encoding a zinc finger protein (Brouwer *et al.*, 2009).

Myotonic dystrophy is an example of a TRED in which the disease is caused by a gain of function mutation, in this case in an RNA molecule. The DM1 transcript produces an RNA with an expanded $(CUG)_n$ tract. Repeat tracts in the single stranded RNA molecule are capable of folding on themselves and forming secondary structures (Napierala and Krzyzosiak, 1997). The proposed mechanism of the disease involves RNA-binding proteins (RNA-BPs) binding to these expanded RNA$(CUG)_n$ structures, which can alter the activity of the RNA-BP. The disease phenotype is caused by this altered protein activity, many of the RNA-BP are involved in pre-mRNA splicing, and aberrant splicing

could produced the phenotypes seen in the disease state (Gatchel and Zoghbi, 2005).

1.2.2.3 Friedreich Ataxia

Expanded TNR arrays in intronic regions of a gene can also be the cause of genetic disease. Friedreich Ataxia is the most common form of inherited ataxia. It is an autosomal recessive disease with characteristic symptoms of limb ataxia, lack of tendon reflexes and weakness in the legs (Campuzano *et al.*, 1996).

This disease is caused by an expanded GAA repeat sequence in the first intron of the gene *FDRA*. It is brought about by an expansion to a repeat length of 200-1700 GAA repeats. A pre-mutation range exists between 33-200 GAA repeats, whilst normal unaffected individuals have less than 32 repeat units in the array at this loci (Gatchel and Zoghbi, 2005). The *FDRA* gene encodes the protein Frataxin, which is involved in the synthesis of haem and iron-sulphur clusters as well as affecting mitochondrial iron levels (Brouwer *et al.*, 2009). The expansion of the TNR in the intronic region of the gene interrupts transcription and leads to a decreased expression of the Frataxin protein (Gatchel and Zoghbi, 2005).

1.2.2.4 Spinocerebellar Ataxias

Two of the spinocerebellar ataxias are caused by expanded TNR sequences in non-coding regions of genes. Spinocerebellar ataxia 8 is caused by a CTG repeat expanded to over 100 repeat units in the 3'-UTR of an untranslated gene (Brouwer *et al.*, 2009). It is a slowly progressive dominant disease that displays characteristic signs of cerebellar ataxia, limb ataxia and spasticity. Magnetic resonance imaging studies show that patients with spinocerebellar ataxia 8 have areas of atrophy on the cerebellum – the area of the brain where learned movements are thought to be stored (Mutsuddi and Rebay, 2005). The molecular mechanism of disease is believed to be due to the fact that the gene overlaps with the 5'-end of the *KLHL1* (Kelch-like 1) gene which is transcribed in the opposite direction. *KLHL1* encodes a protein involved in actin binding. It has been suggested that transcriptional regulation of *KLHL1* is responsible for the disease state (Brouwer *et al.*, 2009).

Spinocerebellar ataxia 12 is also caused by an expanded CAG array in the 5'-non-coding region of a gene *PPP2R2B*. This gene codes for the regulatory subunit B of the protein phosphatase 2A found in the brain. The expansion is believed to influence the transcription as in affected individuals transcript and protein levels are both increased (Brouwer *et al.*, 2009; Gatchel and Zoghbi, 2005).

1.2.3 TNR in coding regions

Many TNR arrays present in transcribed regions of genes are composed of CAG repeat units, which code for the amino acid glutamine, leading to the description of the group as polyglutamine (polyQ) disorders (Gatchel and Zoghbi, 2005). The proteins encoded by these genes are expressed throughout the central nervous system but show distinctive patterns of neurodegeneration in specific neurons (Zoghbi and Orr, 2000).

1.2.3.1 Huntington Disease

Huntington disease is a dominantly inherited polyglutamine disorder. The disease normally manifests in mid-life and is characterized by symptoms of chorea (abnormal involuntary movement), memory problems, and depression with a life expectancy only being 10-15 years following onset (Orr and Zoghbi, 2007).

Huntington disease is caused by an expansion of the CAG TNR array in the gene encoding Huntingtin protein. Unaffected individuals have between 6-34 repeats at this locus, whereas those affected with the disease have over 36 CAG repeats present (Orr and Zoghbi, 2007) with the longest array found to be 121 repeat units long (Zoghbi and Orr, 2000). The exact function of the protein is unknown, though it is known to be involved in pathways of vesicle transport and cytoskeleton anchoring (Li and Li, 2004).

The mechanism by which the mutated protein leads to the disease state is not fully understood. A loss of function of the protein may account for some of the pathology associated with the disease. However, a model of the mutated protein providing a toxic gain of function has also been suggested (Brouwer *et al.*, 2009). Protein inclusion bodies containing the Huntingtin protein have been found to accumulate in brain tissue in patients with the disease, and could possibly produce abnormal protein interactions and aggregation of other proteins into the inclusion (Brouwer *et al.*, 2009).

1.2.3.2 Spinocerebellar Ataxias

CAG repeat arrays are also present in the coding regions of genes implicated in several of the spinocerebellar ataxias: SCA1, 2, 3, 6, 7 and 17 (Orr and Zoghbi, 2007). These diseases are all caused by mutations in different genes, which other than containing an expanded CAG tract share no homology (Brouwer *et al.*, 2009). These diseases again form cellular inclusions of protein suggesting the cell may have a problem processing protein with mutant polyglutamine tracts (Brouwer *et al.*, 2009).

1.3 Trinucleotide repeat instability

The mechanisms by which TNR arrays expand is a topic of much interest. The ability of such arrays to undergo instability events is understood to be influenced by a number of features of the arrays themselves.

1.3.1 *In vitro* thermodynamic stability of TNR sequences

Experimental work *in vitro* has shown that TNR sequences have the ability to form DNA secondary structures of differing thermodynamic stabilities, leading to the suggestion that this capacity to fold may be an important factor in the instability of TNR arrays *in vivo*. TNR arrays have been shown to be able to form into hairpins (Figure 1.1a), tetraplex DNA (Figure 1.1b) and triplex DNA (figure 1.1c) (Sinden *et al.*, 2002).

Numerous *in vitro* studies have found that the TNR sequences associated with disease in humans are susceptible to intrastrand pairing to fold into a hairpin-like structure (Gacy *et al.*, 1995; Mitas *et al.*, 1995a; Mitas *et al.*, 1995b; Petruska *et al.*, 1996; Yu *et al.*, 1995). The structures formed can be referred to as hairpin-like structures as they do not form true hairpins, as palindromic sequences do. The hairpin-like structures formed by TNR are less thermodynamically stable than palindromic hairpins, as they contain a number of mismatches along the length of the hairpin. Comparison between different TNR sequences showed a difference in the thermodynamic stabilities of the hairpin-like structures formed as well (Gacy and McMurray, 1998). CTG TNRs

can form a more thermodynamically stable hairpin-like structure than CAG repeats (Petruska *et al.*, 1996), which can be attributed to the mismatches in the structures. In CTG arrays T•T mispairing would arise, while in CAG arrays A•A mispairing occurs. An A•A mispair is much bulkier than a T•T mispair, and would not fit into the hairpin structure, bulging out (Figure 1.1a). Similar sequence differences were found in the thermodynamic stability of CGG and CCG hairpin-like structures (Mitas, 1997).

Other structures have been suggested to form in TNR arrays, potentially affecting the instability mechanisms they can be subject to. Another of these structures is a tetraplex (4-stranded) DNA which can form in guanine rich sequences, where they form G-quartet DNA (Figure 1.1b). CGG repeat sequences have been suggested to form either hairpin-like structures or G-quartet DNA *in vivo* (Darlow and Leach, 1998; Fojtik *et al.*, 2004; Fojtik and Vorlickova, 2001)

GAA TNR arrays are present in the disease Friedreich ataxia. GAA repeats have been shown to be able to adopt another structure *in vitro,* by folding into a triplex DNA (Figure 1.1c) (Potaman *et al.*, 2004) which consist of a GAA•TTC duplex with normal Watson-Crick pairing, and a third CTT strand paired through non-Watson-Crick pairing. Formation of a triplex structure in GAA repeats may restrict movement of RNA polymerase enzymes through the *FDRA* gene, which could lead to the reduced levels of mRNA characteristic of the condition (Grabczyk and Usdin, 2000).

Unwinding of duplex DNA followed by the strands re-annealing in the wrong positions could produce the slipped strand structure (Figure 1.1d). In a CTG•CAG strain this would produce a CTG hairpin on one strand and a CAG hairpin on the complementary strand. Such structures have been shown to be capable of forming *in vitro* in TNR arrays (Sinden *et al.*, 2002), with the degree to which such structures form being dependent on the length of the array (Pearson and Sinden, 1998).

The ability of TNR arrays associated with disease to form structure *in vitro* is key to the mechanism of instability in these arrays. This is supported by the observation that TNR sequences not associated with disease do not form structure. It is unlikely that any of the non-disease associated TNR would form structure at physiological salt concentrations as any structure would lack the interactions that help to stabilize the structures identified in the structutres formed by disease-associated TNR sequences (Mitas, 1997).

Results detecting secondary structure formation within TNR arrays are all the product of *in vitro* experimentation. Indirect results on the instability of TNR arrays have supported the model of TNR arrays forming hairpin-like sequences *in vivo*.(Zahra *et al.*, 2007). However, no direct evidence for the formation of these structures *in vivo* yet exists.

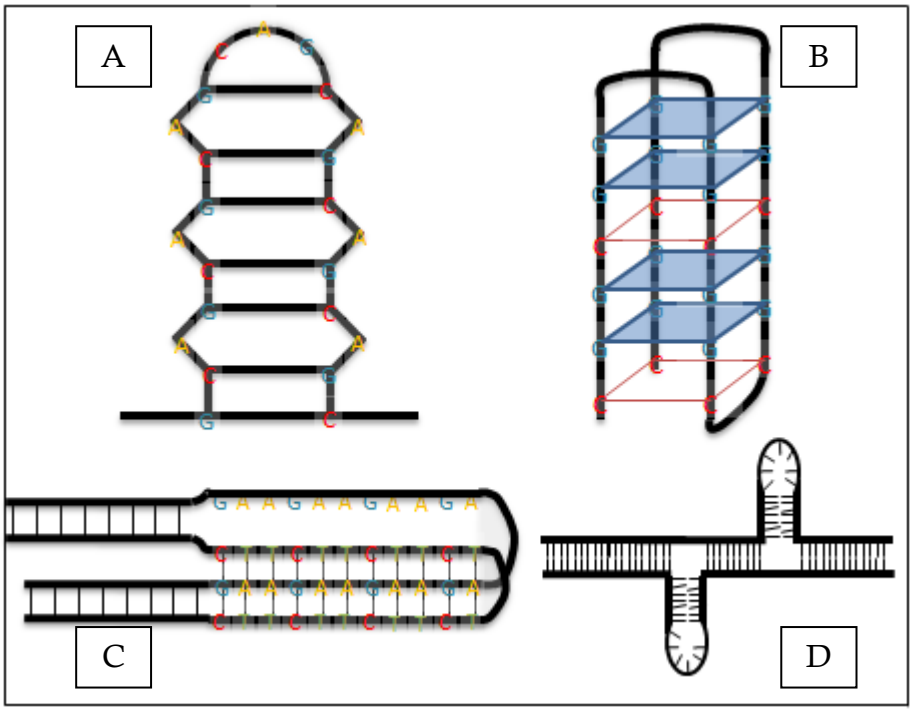

Figure 1.1. DNA secondary structures formed *in vitro* by TNR. Adapted from (Sinden *et al.,* 2002) **A.** Hairpin-like structure. **B.** Tetraplex structure. **C.** Triplex structure. **D.** Slipped strand structure.

1.3.2 *Cis*-elements affect TNR stability

The instability seen in TNR arrays has been shown to be affected by a variety of *cis*-elements *in vivo*. These elements can be both internal and external to the TNR array, including length, orientation and purity of the repeat array, and its chromosomal environment – areas of CpG methylation, or nucleosome bound regions (Cleary and Pearson, 2003).

Deletion instability has been show to be dependent on the length of the TNR array, in a study of CTG•CAG arrays in the *E. coli* chromosome (Zahra *et al.*, 2007). That length of TNR arrays has a direct impact on the instability of the tract is not surprising considering the situation seen in human disease.

Threshold lengths for TREDs provide a block to instability, such that only arrays above that length undergo large expansion events (Pearson *et al.*, 2005; Richards, 2001). In fact, while above the threshold length large expansions predominate in human cells, below threshold a significant number of deletion events occur (Leeflang *et al.*, 1999; Leeflang *et al.*, 1995).

A number of the TREDs show, in the stable TNR arrays of unaffected individuals, interruptions to the repeat array, while the expanded arrays found in those affected by these diseases lack such interruptions (Rolfsmeier and Lahue, 2000). Spinocerebellar ataxia 1 is caused by a CAG TNR array of 40 repeat units or more, at which length it becomes unstable. In those unaffected by the disease the CAG array has a midstream CTG interruption. In fact, a TNR array of 39 CAG repeats with one CTG interruption the TNR array remains stable, while in those with an array containing 40 CAG repeats with no interruption the TNR becomes unstable (Chong *et al.*, 1995).

The orientation of TNR arrays in the chromosome, relative to the direction of replication, has been shown to be an important *cis*-determinant affecting instability in TNR arrays. Experimental work in yeast and bacterial cells harbouring CTG•CAG TNR arrays have shown that deletion instability is higher in cells which carry the CTG repeat array on the lagging strand template, while some studies suggest that expansion instability is higher in the orientation in which CTG repeats are present on the newly synthesised lagging strand (Freudenreich *et al.*, 1997; Kang *et al.*, 1995a; Miret *et al.*, 1998). The orientation dependence of TNR instability is believed to be due to the ability of these

sequences to form hairpin-like structures, as shown *in vitro.* CTG TNRs form more thermodynamically stable hairpin-like structures, which if present on the lagging strand template could form in the transiently single stranded region of the discontinuous lagging strand. *In vivo* experiments on chromosomal TNR arrays support this model of hairpin formation during replication leading to instability (Zahra *et al.*, 2007). Experiments in human cell lines have identified potential replication initiation regions near loci associated with TREDs. These potential initiation regions would allow replication in the orientation predicted to give the instability seen in the disease states. They have also shown that the distance of the TNR array from a replication initiation region also plays a role in the instability at the array (Cleary *et al.*, 2002; Nenguke *et al.*, 2003).

Cis-elements external to the TNR array have also been implicated in affecting instability of the repeats. CpG methylation may stabilize TNR arrays, as demonstrated in the CGG arrays of fragile X patients in which the expanded arrays remain that remain stable in somatic cells when methylated, but can become unstable when unmethylated (Cleary and Pearson, 2003). In higher eukaryotes flanking regions of DNA have been shown to be important in the instability of TNR arrays. The CTG array in the DM1 locus associated with Myotonic Dystrophy I has been shown to be flanked by two CTCF binding sites (Filippova *et al.*, 2001). CTCF is a zinc-finger DNA binding protein which has roles in genome imprinting and chromatin insulation. A study disrupting CTCF binding sites at an SCA7 locus in transgenic mice found increased TNR

instability both in somatic tissue and through vertical transmission (Libby *et al.*, 2008).

The general model by which TNR instability is thought to occur *in vivo* involves strand slippage at the TNR array. Such slippage events could be stabilized through the formation of secondary structure within the slipped sequence. The differing thermodynamic stabilities of structures formed by TNR sequences would explain the difference in instability associated with each TNR sequence. The process of DNA synthesis is a likely mechanism to allow such slippage events to occur. DNA synthesis occurs in the cellular processes of replication, recombination and DNA repair.

The generalized pathways leading to deletion or expansion events describes deletion events occurring when the slipped strand occurred on the template strand of DNA synthesis (Figure 1.2a), whereas expansions would arise from a slipped strand on the newly synthesized strand (Figure 1.2b) (Bichara *et al.*, 2006)

Throughout this thesis models have been drawn in the same way, with CTG TNR arrays coloured green and CAG TNR arrays coloured orange.

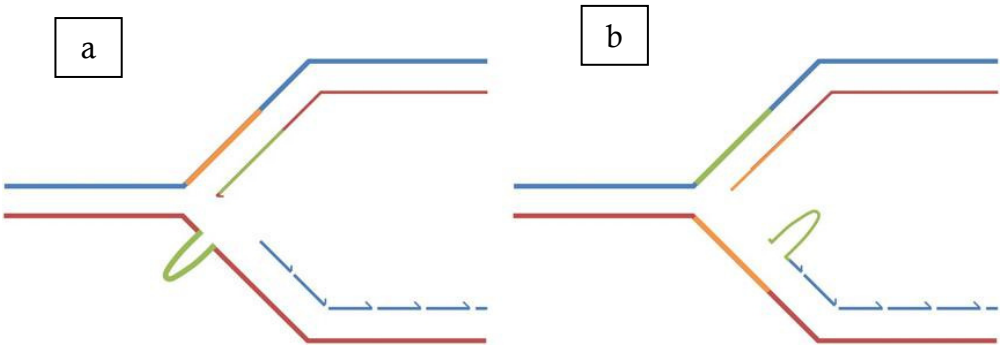

Figure 1.2. Slipped strand model of TNR instability. a. Slipped strand on template strand would lead to a deletion event in the TNR array, following a subsequent round of replication. **b.** Slipped strand on the newly synthesized strand would produce an expansion following a subsequent round of replication.

1.4 Cellular processes implicated in TNR instability

1.4.1 *E. coli* DNA replication

Normal growth and division of cells requires regular and coordinated duplication of the genome. The vast amount of information held in the DNA is duplicated through the process of DNA replication. *E. coli* is a classic system for studying these processes, and as many of the players involved are conserved throughout different organisms, the mechanisms involved may be generalized (Yao and O'Donnell, 2009).

Replication of the chromosome is described as a semi-discontinuous process as the synthesis of each strand is accomplished differently. The leading strand is replicated as a continuous, uninterrupted chain, while the lagging strand is synthesised in short chains of less than 2kb called Okazaki fragments (Langston and O'Donnell, 2006).

Chromosomal replication is initiated at origin sites; in *E. coli* there is a single origin site, the approximately 250bp *oriC* (Mott and Berger, 2007). Initiation of replication is controlled by the protein DnaA. DnaA can bind ATP and this complex is able to bind the origin region, causing the melting of the duplex DNA in the DNA unwinding element (DUE) region of the *oriC* (Bramhill and Kornberg, 1988). Once this region of the chromosome has melted to produce single stranded DNA, the DnaA protein recruits the replicative helicase DnaB from the DnaB-DnaC complex (Marszalek and Kaguni, 1994). DnaB protein acts in a homohexameric protein complex that encircles the lagging strand template and travels to the apex of the replication forks where it utilizes energy from ATP to translocate along the DNA separating the parental duplex ahead of it (Figure 1.3) (Yao and O'Donnell, 2009). The group of proteins acting together to advance the replication fork and synthesize new DNA is called the replisome. The first stage of the DNA synthesis process is production of an RNA primer, which is accomplished by the replisome primase DnaG (Langston and O'Donnell, 2006). The synthesis of the new DNA strands is achieved by the DNA polymerase III holoenzyme. This large protein complex consists of three Pol III core polymerase (McInerney *et al.*, 2007; Reyes-Lamothe *et al.*), two β sliding clamp processivity factors and the clamp loader (Figure 1.3) (Johnson and O'Donnell, 2005). The Pol III core polymerases are made up of the αεθ subunits. The α subunit, encoded by the *dnaE* gene, is the polymerase. The ε subunit is encoded by the *dnaQ* gene and is a 3'-5' proofreading exonuclease (Johnson and O'Donnell, 2005). The DnaG primase remains bound to the RNA

primer until it is replaced by the Pol III core polymerase, allowing it to cycle back to prime another Okazaki fragment (Langston and O'Donnell, 2006). The polymerase III holoenzyme acts to allow replication of the DNA in a highly processive fashion, a normal *E. coli* replication fork moves at around 1kb per second (Langston and O'Donnell, 2006). The advancing replisome permits continuous extension of the nascent leading strand. On the lagging strand, the polymerase extends an Okazaki fragment. However as the two polymerases of the replisome are connected via the clamp loader, a loop is generated in the lagging strand as DNA is extended in the opposite direction (Figure 1.3) (Johnson and O'Donnell, 2005).

Once replication of an Okazaki fragment has completed the Pol III core polymerase must then detach from the newly synthesized Okazaki fragment. Two models have been suggested for how this may occur; in the first, dissociation of the Pol III core polymerase is achieved through intramolecular signalling, when the complex encounters the 5'-terminus of the preceding Okazaki fragment. Normally the Pol III core polymerase is strongly bound to the DNA with a dissociation half life of 5 minutes, however after collision with the previous Okazaki fragment this interaction is disturbed so dissociation half life falls to only 1 second (Stukenberg *et al.*, 1994). This allows the polymerase to cycle back to a new RNA primer, but leaves the β clamp attached to the DNA. This remaining β clamp may then play a role in processing of the Okazaki fragment, removing the RNA primer and ligating Okazaki fragments together to form a continuous lagging strand, as the β clamp is able to bind other proteins

useful for such activities such as DNA Polymerase I and DNA ligase (Johnson and O'Donnell, 2005). In the second proposed model the Pol III polymerase is released from the β clamp before the Okazaki fragment has finished (Yao and O'Donnell, 2009).

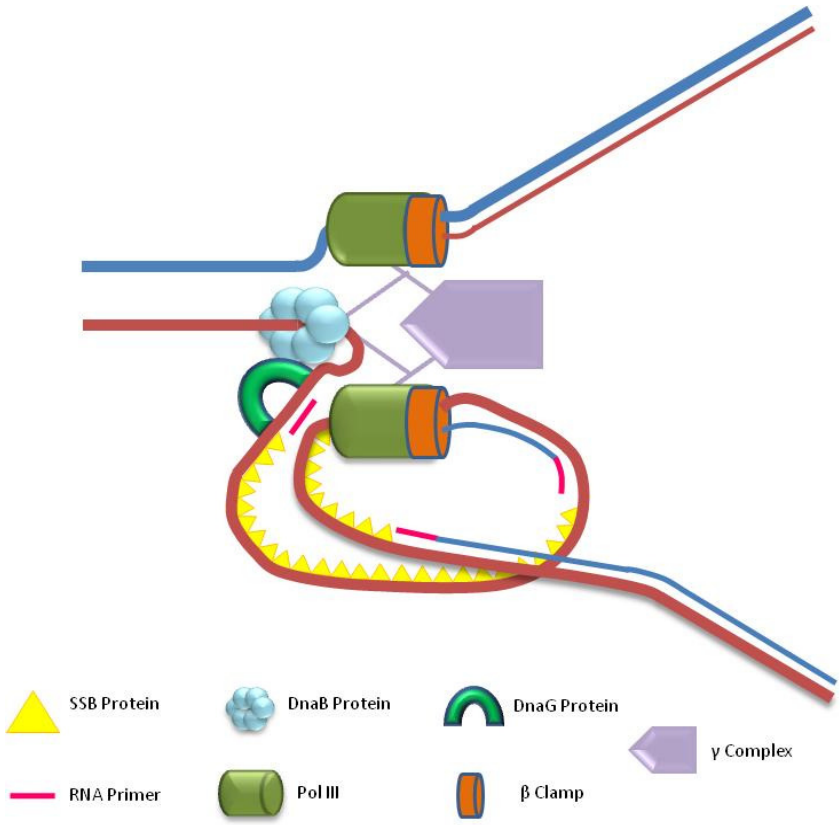

Figure 1.3. The *E. coli* replication fork. (Adapted from (Yao and O'Donnell, 2009). Though three DNA polymerases may be present only two have been shown for simplicity.

In Eukaryotic cells many of the components of the replisome are comparable to those found in the prokaryotic replisome, though additional factors are involved. In eukaryotic cells rather than dnaG RNA primers are generated by a primase enzyme complexed witht he protein polα, also instead of the βclamp eukaryotic cells utilize PCNA (proliferating cell nuclear antigen). The replicative helicase in

eukaryotes is a heterohexameric complex of MCM proteins (MCM2-7), rather than the DnaB homohexameric helicase found in *E.coli*. Eukaryotes also differ in that they utilize two different polymerases, polε and polδ (Yao and O'Donnell, 2009).

1.4.2 Replication fork blockage

The process of replication does not go unchallenged during the cell cycle as movement of the fork through the duplex DNA can be hindered in a number of different ways.

The moving replication fork can run into proteins tightly bound to the DNA in the duplex ahead of the replication fork, such as RNA polymerases transcribing open reading frames within the DNA (Atkinson and McGlynn, 2009). RNA polymerase enzymes move along the template DNA in either direction depending on the gene they are transcribing, so collision with the replisome is inevitable (Rudolph *et al.*, 2007). If a collisions were head on, when both complexes are moving in opposite directions towards each other, stalling of both could be caused by the physical collision of the two complexes. Alternatively studies have suggested DNA topology could be the cause of this stalling. Both the transcription and replication processes produce positive supercoils ahead of the moving protein complex involved, if these were moving toward each other it would produce a great torsional stress on the region of DNA between, enough to block movement of either complex (French, 1992).

Replication forks encounter pre-programmed blocks at the Tus/*ter* sites in the *E. coli* chromosome. These sites act as a replication fork trap for forks in which replication has passed the normal termination point, as a fork is able to move through it in one direction but not the other (Hill *et al.*, 1987). The *ter* site is bound by Tus protein producing the so-called Tus/*Ter* complex, which prevents DnaB protein action, arresting the fork (Mulcair *et al.*, 2006).

Lesions in the DNA may also restrict movement of the replication fork. Lesions are sites of damage to the DNA which can alter the shape of the duplex strands, potentially preventing replication fork progression (Courcelle *et al.*, 2003). This may be beneficial to the cell in helping to avoid replication of damaged DNA. Alternatively DNA replication may be able to bypass the lesion, leaving it to be repaired post-replication (Heller and Marians, 2006a). Whether lesions lead to replication fork stalling or a lesion bypass mechanism, is reflective of the difficulty of determining the degree to which replication forks actually stall *in vivo* as there are multiple repair pathways in the cell to deal with such structures (Atkinson and McGlynn, 2009).

When the replication fork encounters a challenge it can stall, no longer moving forward through the unreplicated DNA, producing a blocked fork structure. Studies in which replication restart proteins have been mutated have provided some evidence to the extent at which this stalling occurs *in vivo.* DnaC protein, which is believed to load the replicative helicase DnaB onto arrested forks, was inactivated (using a temperature-sensitive mutation), resulting in

approximately 18% of cells failing to replicate their chromosome in one round of replication (Maisnier-Patin *et al.*, 2001). Further studies have suggested an even greater extent of replication fork blockage in *E. coli*. PriA and PriC protein have been shown to be required to restart different species of blocked fork structure (Heller and Marians, 2006b). Double mutants of these proteins are inviable, suggesting the presence of at least one is essential for survival of the cell suggesting at least one fork needs to be reinitiated every replication cycle (Sandler, 2000).

Blocked fork structures may experience a disassembly of the replisome. The stability of the replisome proteins in an arrested fork has been studied *in vivo* in *E. coli* using reversible blocks to replication. It was found that once the block was removed resumption of replication was a rapid process, even after two hours of arrest, suggesting that the replisome proteins at least remain in the vicinity of an arrested fork (Possoz *et al.*, 2006).

The structure of the stalled fork produced may well be dependent on the nature of the stall itself. One can imagine three different situations that could lead to arrest of replication, one which is caused by a block on both the leading and lagging strand of the duplex DNA, one caused by a block on the leading strand only and one from a block in the lagging strand replication. Replication forks which have arrested due to a blockage extending over both strands have been shown to contain a single stranded region on the lagging strand template in the order of tens of bases long, and a small single stranded region on the leading strand only a few bases long (reviewed in (Atkinson and McGlynn, 2009)).

Forks which encounter a block to replication on the lagging strand template have been shown to continue leading strand synthesis producing more single stranded DNA on the lagging strand template. This would allow further priming of the lagging strand template away from the blocking lesion and cycling of the polymerase to this new primer, thereby skipping an Okazaki fragment containing the lesion, which can then be repaired later (Langston and O'Donnell, 2006). Replication forks blocked by a lesion specifically on the leading strand prove slightly more problematic to the cell. A number of different ways have been suggested in which a challenge such as this could be overcome. Arrest of the leading strand polymerase could lead to an uncoupling of the leading and lagging strand replication, if lagging strand replication continues a single stranded gap would appear on the leading strand template. *In vivo* plasmid studies in *E. coli* have shown in these systems a gap of at least 1kb is produced (Pages and Fuchs, 2003).

A model of how leading strand gaps could be resolved involves template switching. The newly synthesized leading and lagging strands may regress to anneal with each other forming a Holliday junction. In this structure the single stranded region of newly synthesized lagging strand could then be used as a template for synthesis of the nascent leading strand. Restructuring of the replication fork after such an event would produce a substrate for PriA mediated reactivation of the fork with the lesion by-passed (Heller and Marians, 2006a).

1.4.3 Reinitiation of replication

The PriA and PriC proteins are key players in the restart of arrested replication forks, differing in the structures on which they work (Heller and Marians, 2006b). Unlike normal replication initiation that occurs at the origin sequence, reinitiation of replication must have the ability to occur anywhere in the genome, independent of sequence as damage causing arrest is not limited to specific genomic loci. PriA protein is a helicase belonging to the superfamily 2 class of helicases in *E. coli* (Hall and Matson, 1999). PriA binding to the DNA at arrested replication forks leads to the recruitment of a number of proteins collectively termed the primosome (PriB, DnaT, DnaB, DnaG), which act to allow reinitiation of replication and assembly of a functional replisome (Sandler and Marians, 2000). PriA acts at arrested forks in which the cause of the arrest has not generated a leading strand template gap. The ability of the protein to recruit the other primosomal proteins depends on the presence of a 3'-end of a newly synthesized leading strand (Heller and Marians, 2005a). In forks arrested as species containing a single stranded gap on the leading strand template, it is not the helicase PriA that is responsible for the restart of replication, but the protein PriC. Under these circumstances PriC acts to restart replication by allowing the helicase Rep to unwind any nascent lagging strand at the fork, as the presence of such would prevent the PriC protein from loading the replicative helicase DnaB, required for replication restart (Heller and Marians, 2005b).

1.4.4 Replication fork reversal

Replication fork structures in which direct restart by the PriA or PriC proteins cannot be accomplished, can be restructured by replication fork reversal (RFR) to enable removal of the site of arrest and formation of a structure capable of reinitiation. The idea of forks being able to reverse was first proposed in 1974 (Hotchkiss, 1974) but received little attention until the 1990s.

Figure 1.4 shows a model for the processing of a reversed replication fork (Seigneur *et al.*, 1998). The tetrameric RuvA protein complex binds the arrested replication fork and promotes the assembly of a hexameric complex of RuvB protein (Figure 1.4a). RuvB can then pull the parental strands through the RuvA complex peeling off the newly synthesized strands, which can then anneal and form a Holliday junction (HJ). This HJ structure allows the assembly of a second RuvB complex, which further drives reversal of the replication fork (Figure 1.4b) (Baharoglu *et al.*, 2006). A feature of RFR is that the process generates a double stranded end in the DNA without the need for a chromosomal breakage event (Michel *et al.*, 2004). This end is a substrate in *E. coli* for the RecBCD complex (Figure 1.4c), a heterotrimeric complex which has both helicase and exonuclease activity. As the complex translocates along the DNA it can degrade it until it either reaches the HJ (Figure 1.4d) thereby restoring the fork DNA structure, or it encounters a specific DNA sequence called *chi* (Figure 1.4e) at which the process of homologous recombination (HR) is initiated (see Chapter 1.4.5). At *chi* sites the RecBCD complex loads the protein RecA onto the DNA, forming a RecA filament (Anderson and

Kowalczykowski, 1997). This filament is capable of strand invasion into DNA containing homologous sequence establishing a HJ. The RuvABC proteins can then migrate and then cleave the HJ to form a structure capable of replication restart by the PriA protein (Michel *et al.*, 2004). How arrested replication forks enter this pathway has been investigated in strains deficient in specific replication proteins to produce arrest. Several mechanisms by which an arrested replication fork can be reversed, converted from a three-armed structure to a four-armed structure, are believed to exist *in vivo* (Michel *et al.*, 2004).

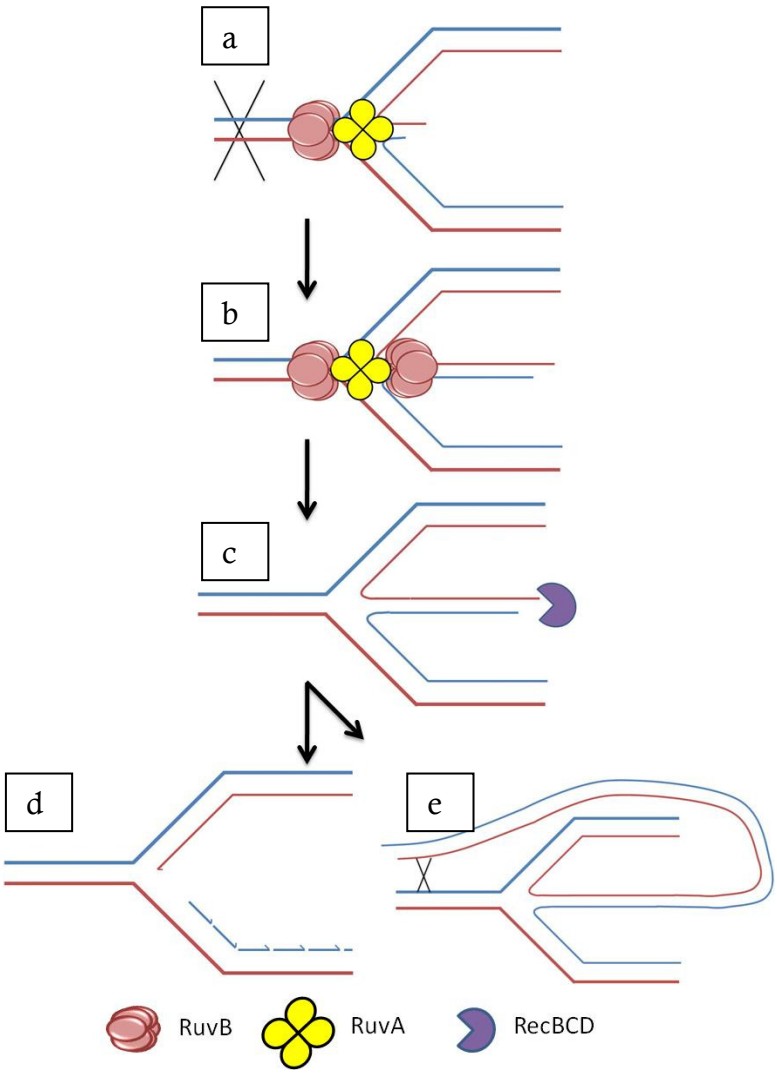

Figure 1.4: Model of processing of reversed replication fork. Adapted from (Seigneur *et al.*, 1998). Stalled replication forks bound by RuvAB (**a**) can separate daughter strands from their templates, reversing the replication fork (**b**). This produces a dsDNA end, a substrate for the RecBCD nuclease (**c**) which will digest the daughter strand back to the apex of the fork (**d**) or to a *chi* site at which HR will be initiated (**e**).

One model of RFR does not involve any enzyme action but relies on the driving force of supercoiled DNA. The local unwinding of duplex DNA at a replication fork generates positive supercoils ahead of the fork apex. If the fork progresses and this supercoiled DNA is not relaxed a great torsional strain will

develop in the DNA. Experiments conducted using plasmid DNA intercalating agents to mimic the strand separation achieved by the replication fork found that upon analysis by scanning force microscopy the torsional strain that developed had been relieved by RFR (Postow *et al.*, 2001). However, as this study was conducted using a plasmid based assay the question remains as to how closely it reflects the ability of this to occur in chromosomal DNA *in vivo*. The reactions described by this model would be dependent on the equilibrium between how quickly the replisome could disassemble allowing the RFR to occur, and how quickly the cellular topoisomerase enzymes could act to relieve the torsional strain negating the need for the RFR mechanism (Atkinson and McGlynn, 2009).

Inactivation of the replicative helicase DnaB causes arrest of the replication fork as the duplex DNA ahead of the fork cannot be unwound. Experimental work conducted in an *dnaBts* mutant strain found that when replication was arrested by growing at the non-permissive temperature RFR was dependent on RecA protein (Seigneur *et al.*, 2000). A model of how the DnaB arrested fork progressed to RFR via a RecA dependent pathway was developed. RecA protein was proposed to bind the single stranded lagging strand template at the arrested fork. The RecA filament then allows strand invasion into the homologous leading strand displacing the nascent leading strand, allowing it to pair with the nascent lagging strand, producing a Holliday junction (Seigneur *et al.*, 2000).

The above RecA dependent pathway identified in *dnaBts* cells is unique in that experimental work in which fork arrest was caused through inactivation of other replication proteins found RFR to be RecA independent (Flores *et al.*, 2005). Studies into replication forks arrested by inactivation of polymerase III holoenzyme proteins DnaE (the catalytic subunit), DnaN (the processivity factor) or HolD (a clamp loader protein) lead to the development of an alternative RecA independent model of RFR (Figure 1.5) (Flores *et al.*, 2005). It was shown that in these mutant strains, though RecA was not required for RFR, UvrD helicase was required (Flores *et al.*, 2004). UvrD is a 3'-5' helicase shown to be able to clear RecA protein from DNA (Centore *et al.*, 2009). The model developed suggests that RecA protein in these mutant strains is actually causing a block to RFR, while UvrD helicase will clear RecA filament from the lagging strand template allowing RFR to take place. This hypothesis is supported by the observation that the requirement for UvrD helicase is lost in strains carrying a mutation in the *recFOR* or *recA* genes. The RecFOR complex acts in the cell to load RecA protein onto ssDNA (Morimatsu and Kowalczykowski, 2003). Cells unable to load RecA would not require the UvrD protein for this pathway of RFR to proceed (Flores *et al.*, 2005).

The different pathways of RFR may reflect the importance of this process to the cell, with different pathways being involved at different stalled fork structures. The use of replication protein mutants in studying these RFR pathways may complicate the interpretation, and the multiple different pathways

suggested may be indicative of the difficulties of predicting the wild type situation from mutant studies (Atkinson and McGlynn, 2009).

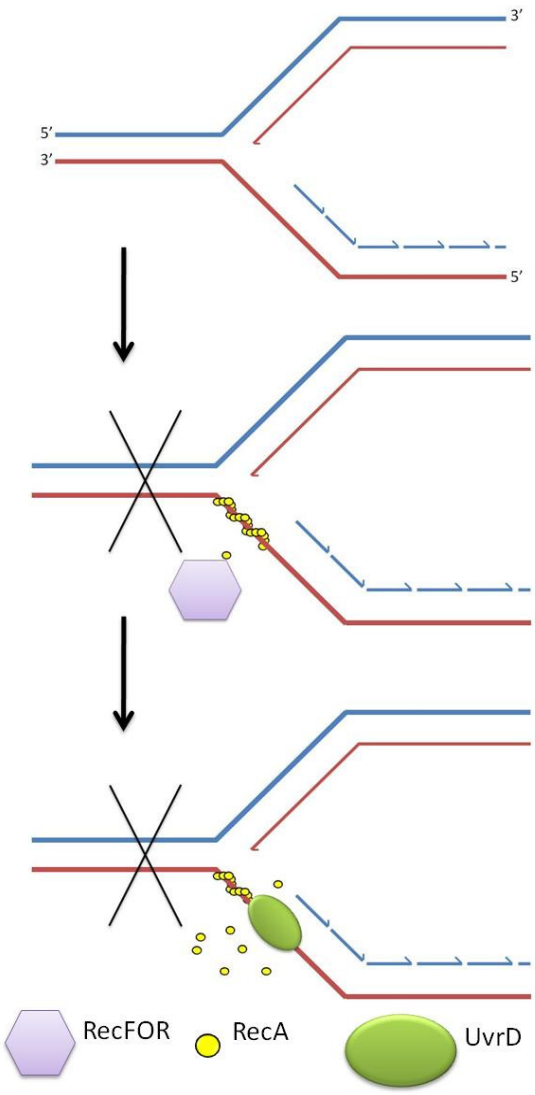

Figure 1.5 Stalled Replication Fork Processing. Adapted from (Flores *et al.*, 2005). RecFOR loads RecA protein onto the lagging strand template of an arrested replication fork. The RecA filament acts to prevent RFR in this pathway, but can be cleared by UvrD helicase.

1.4.5 Double strand break repair

Breaks to both strands of DNA, double strand breaks (DSB), are toxic to the cell. If they persist they can lead to death in prokaryotes and can lead to cancer and major genome rearrangements in human cells if repaired incorrectly. Though potentially dangerous some organisms can induce DSB in their own DNA, for example crossing over in the process of meiosis (Gerton and Hawley, 2005). DSB can be repaired by one of two mechanisms in many organisms; Non-Homologous End Joining (NHEJ) involves the ligation of the two broken ends together, but it is a mechanism absent in *E. coli* (see review (Lieber *et al.*, 2003)). Homologous recombination (HR) is the other process by which DSB can be repaired (Figure 1.6). This evolutionarily conserved mechanism is the pathway utilized in *E. coli* (Cromie *et al.*, 2001). HR is also involved in the rescue of arrested replication forks by replication fork reversal (Michel, 2000). HR utilizes an intact copy of the broken region as a template for accurate repair. There are four steps involved in the repair of a DSB by HR, these are: initiation, synapsis (homologous pairing and strand invasion), branch migration and resolution of joint molecules.

Initiation of HR is dependent on the RecBCD protein complex. RecBCD has both helicase and nuclease activities. It binds to the double strand end and translocates along it, digesting the DNA until it reaches a specific 8bp sequence called *chi* (5'-GCTGGTGG-3') (Figure 1.6b). This *chi* sequence alters the activity of the RecBCD enzyme, attenuating the nuclease activity acting on the strand with a 3'- terminal, which results in the formation of a ssDNA 3'-

overhang (Dillingham and Kowalczykowski, 2008; Dixon and Kowalczykowski, 1993). RecBCD is also responsible for the loading of RecA protein onto this 3'- overhang, by removing the SSB protein present (Anderson and Kowalczykowski, 1997). This action is essential as SSB competes for ssDNA much more efficiently than RecA (Figure 1.6c).

The synapsis process involves the RecA filament formed on the ssDNA 3'- overhang finding homologous sequence on the intact chromosome and extending into it. This strand invasion causes the ejection of the existing paired strand in the formation of a D-loop. D-loops are recognized by the protein PriA which can lead to the recruitment of the replication machinery, establishing a replication fork at the site to allow DNA synthesis using the intact chromosome as a template, repairing the DSB (Figure 1.6d) (Cadman and McGlynn, 2004).

The synapsis stage of DSB repair leads to the production of a joint molecule: the two chromosomes are physically linked, so must be separated for the cell to survive. To accomplish this the D-loop is converted into a four-way Holliday Junction (HJ) (Cromie *et al.*, 2001). HJs are processed by the proteins RuvAB which allow branch migration of the junction, until it is cleaved by RuvC protein, resolving the joint molecule (Figure 1.6e) (Kaplan and O'Donnell, 2006).

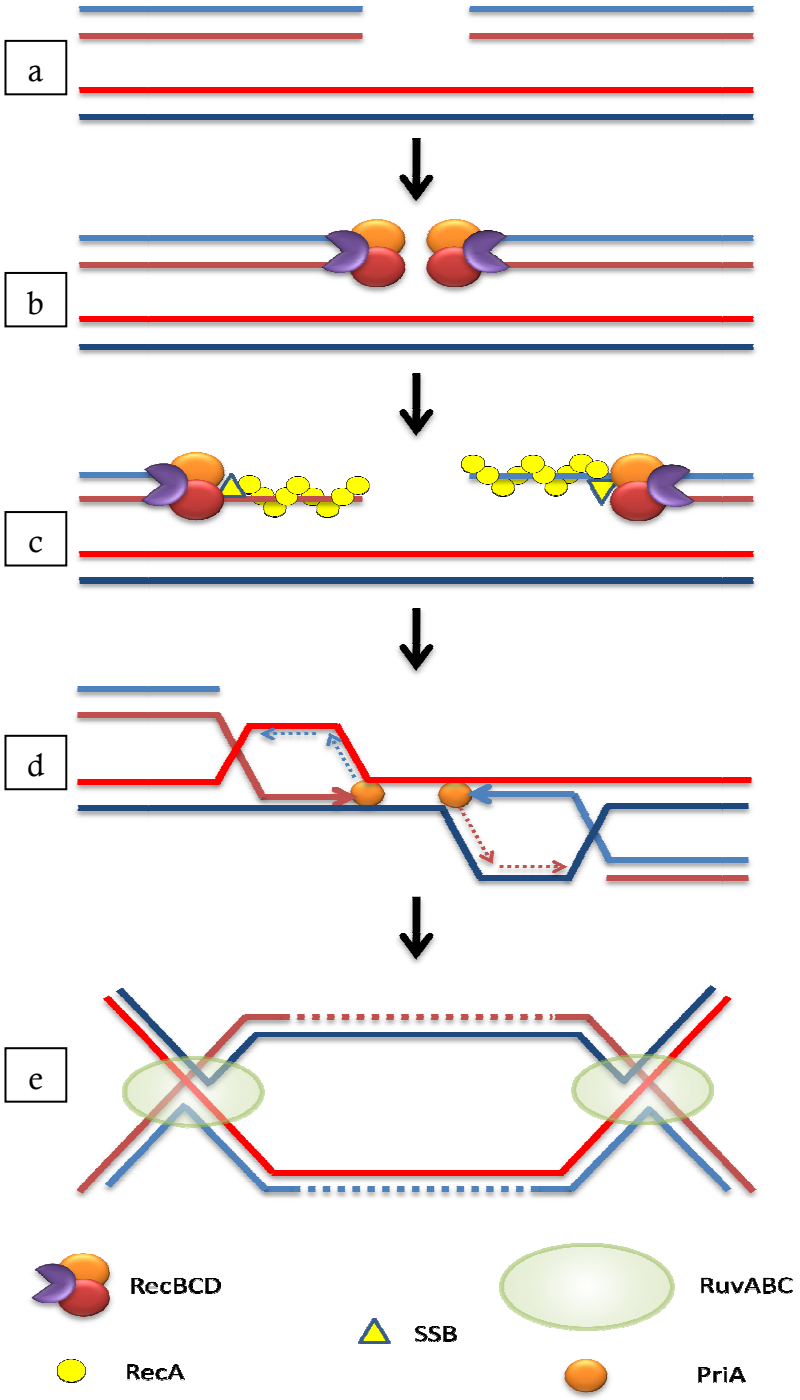

Figure 1.6. Double strand break repair in *E. coli.* **a.** Double strand break in DNA can be repair by HR. **b.** Break digested by RecBCD complex. **c. RecBCD displaces SSB protein loading** RecA onto 3'-ends allowing strand invasion. **d.** PriA establishes replication fork at D-loop. **e.** Double HJ resolved by RuvABC complex.

1.4.6 Gap repair

HR is also utilized in the cell to repair single stranded gaps in DNA by the RecFOR-mediated gap repair pathway (Figure 1.7) (Morimatsu and Kowalczykowski, 2003). The pathway involves a number of the recombination proteins including RecF, RecO, RecR, RecA, RecQ, RecJ, RecG, RuvA, RuvB, RuvC and SSB proteins (Kowalczykowski *et al.*, 1994).

The repair of gapped DNA occurs through a process that is less well defined than repair of double strand breaks to DNA. ssDNA present in gapped DNA structures would be coated by SSB protein (Figure 1.7a), presenting a blockage to RecA filament production on the ssDNA. The RecFOR complex is employed to overcome this problem. RecF binds the gapped DNA at the ssDNA-dsDNA junction (Figure 1.7b), and can complex with RecR, which itself can interact with RecO – suggesting RecR acts as a bridge in the RecFOR complex. The RecFOR complex can then displace SSB protein from the ssDNA, promoting nucleation of the RecA filament (Figure 1.7c) (Morimatsu and Kowalczykowski, 2003). The RecA filament induces pairing of the damaged DNA with the homologous sequence on the intact chromosome, and strand exchange (Figure 1.7d). Finally, resolution of the joint molecule following DNA synthesis to fill the gap, would be accomplished by the RecG or RuvABC proteins (Figure 1.7e) (Morimatsu and Kowalczykowski, 2003).

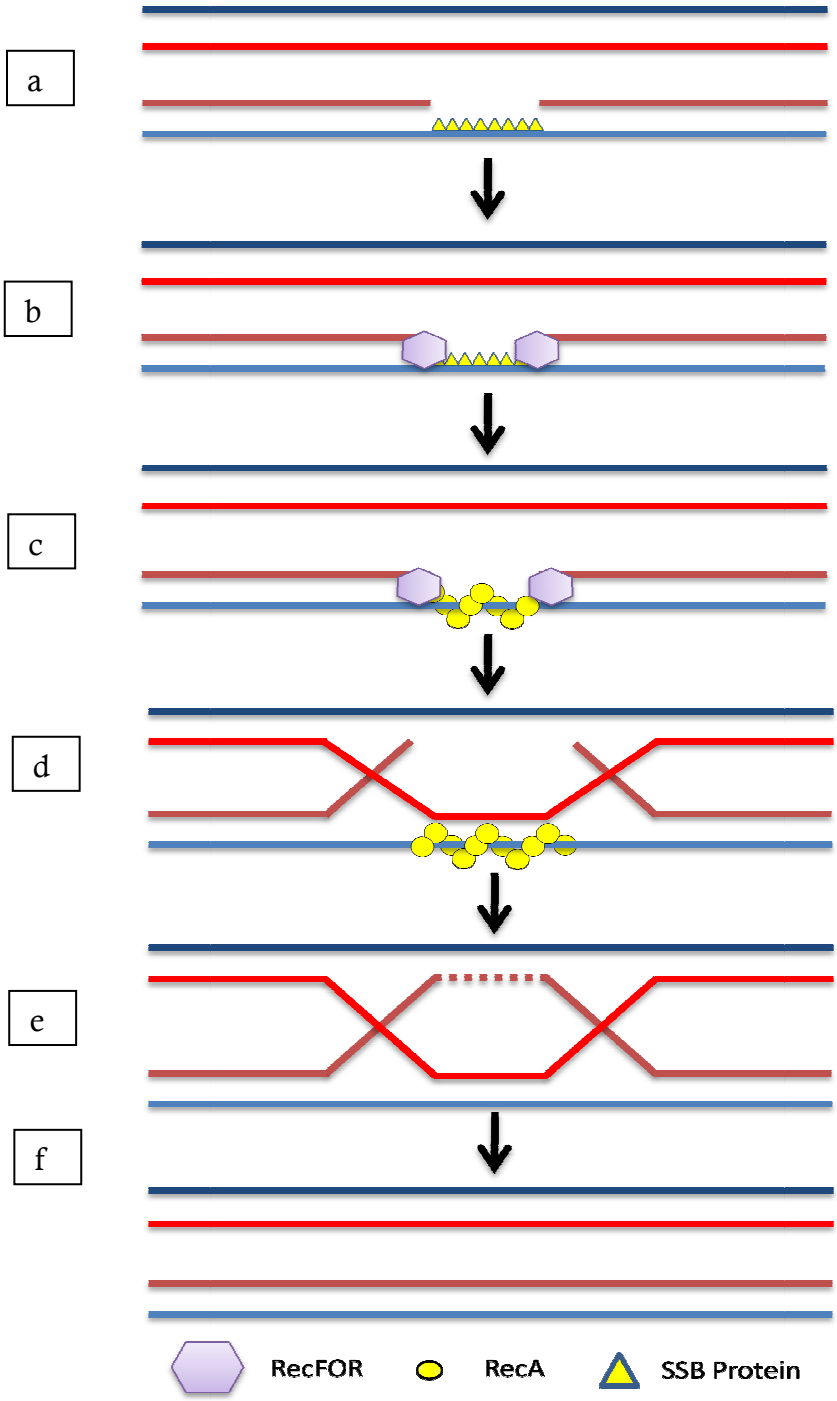

Figure 1.7. Gap repair by HR. Adapted from (Kuzminov, 1999) **a.** ssDNA in a gap coated by SSB protein. **b.** RecFOR complex binds at the ds-ssDNA junction. **c.** RecFOR enables replacement of SSB with RecA protein. **d.** Strand invasion of the homologous sequence on intact chromosome. **e.** DNA synthesis fills gap. **f.** Resolution of joint molecule.

1.5 Models of *in vivo* TNR Instability

1.5.1 DNA replication and TNR instability

Replication of the genome is an attractive candidate pathway for the instability of TNR arrays. The role of this pathway in TNR instability is supported by a number of observations from experimental work conducted on instability of plasmid based TNR arrays in *E. coli*. The growth phase of the host *E. coli* cell and the mode of replication of the plasmids involved in the experiment were shown to have an impact on the instability TNR arrays in plasmid DNA (Bowater *et al.*, 1996; Schumacher *et al.*, 1998), which imply an effect of replication on the instability of TNR arrays. Additionally the observation that TNR instability is dependent on orientation of the repeat array relative to the origin of replication (Kang *et al.*, 1995a) supports this idea. Mutations in genes encoding proteins involved in the replication complexes, such as the PolIII holoenzyme, have been shown to destabilize the TNR array, further supporting replication as a major source of TNR instability (Iyer *et al.*, 2000; Morag *et al.*, 1999; Zahra *et al.*, 2007).

A replication based model in which structure formation was the key factor in instability was developed using the fact that lagging strand synthesis is discontinuous. During replication, a region of the lagging strand template is transiently single stranded. When CTG TNRs were present on the lagging strand more deletion instability was detected than when CAG TNR were present in this region (Kang *et al.*, 1995a). That CTG TNR sequences form more stable

secondary structures *in vitro* then CAG TNR sequences (Gacy and McMurray, 1998), could provide an explanation for the orientation dependence observed. CTG hairpins would likely persist longer than CAG structures formed as CAG structures may form unstructured, looped out DNA, rather than a stable hairpin (Pearson *et al.*, 2002). Following a subsequent round of replication structures that persisted on the lagging strand template would lead to a deletion event in the TNR sequence. Kang and colleagues also saw an increased level of expansion instability when the CTG sequence was present on the leading strand template (Kang *et al.*, 1995a). As expansion instability is believed to arise from slipped strands on the nascent strand of the DNA synthesis process, this orientation dependence may arise from CTG hairpins forming in slipped strands of Okazaki fragments (Bichara *et al.*, 2006). However, the observation of orientation dependence on TNR instability was not found in all studies of TNR instability in plasmids (Schumacher *et al.*, 1998).

Further evidence for the involvement of DNA replication in the instability pathway of TNR comes from the yeast *Saccharomyces cerevisiae*. Work conducted in this model organism showed that *rad27* mutants displayed elevated levels of expansion instability in CAG and CCG TNR tract (Freudenreich *et al.*, 1998; Schweitzer and Livingston, 1998; Spiro *et al.*, 1999). The *RAD27* gene encodes a protein named FEN1 (flap endonuclease 1). FEN1 is a nuclease, which acts in the cell to process 5'- unannealed flaps generated in Okazaki fragment maturation (Kovtun and McMurray, 2008). If a TNR sequence is present in the 5'-flap of an Okazaki fragment it may be able to form into a

hairpin-like secondary structure, an intermediate for instability (Figure 1.8a). Hairpin structures, however, are not suitable substrates for FEN1. FEN1 was shown to compete with the enzyme DNA ligase1. FEN1 prevents expansions, while DNA ligase1 promotes them (Liu and Bambara, 2003). One suggested model to explain the action of FEN1 proposed a flap containing a hairpin structure formed from a TNR array may undergo a further slippage and misalignment event converting the hairpin into a bubble structure with a small region at the 5'-end of the flap annealed to it complementary strand (Figure 1.8b). Large bubble structures may be in equilibrium with smaller bubble structures, with larger regions of duplex DNA at the 5'-end. The gapped DNA in these may then be sealed by DNA ligase1, producing an expansion product. Alternatively large bubbles could be converted to a flap structure (again in an equilibrium reaction) which could be excised by FEN1, which following repair of the gap would produce a non-expanded product (Liu and Bambara, 2003).

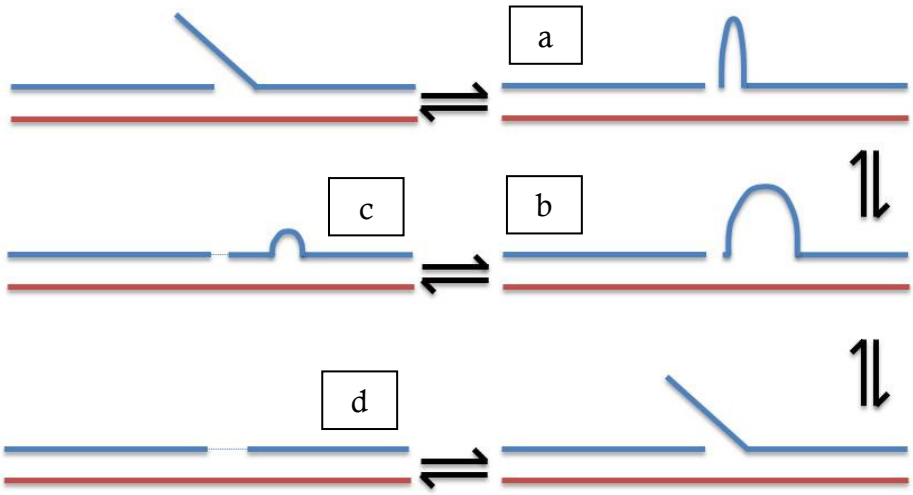

Figure 1.9. Model of FEN1 / DNA Ligase1 pathway of TNR expansion. (Adapted from (Liu and Bambara, 2003). **a.** Okazaki flap containing TNR may form hairpin structures, which are unsuitable substrates for FEN1. **b.** Slippage of the hairpin structure could form a large bubble, **c.** which may further slip into a substrate accessible to DNA Ligase1 – producing an expansion. **d.** Alternatively, the bubble could convert to a flap, accessible to FEN1 producing a non-expanded product.

1.5.2 Replication Fork Reversal (RFR) and TNR instability

Another model for the instability of TNR arrays is based on the idea of replication fork arrest and subsequent reversal. 2D gel electrophoretic studies on replication of repetitive DNA have shown that CCG•CGG and CTG•CAG TNR arrays are able to arrest replication forks i*n vivo* in bacterial plasmid based assays (Samadashwily *et al.*, 1997), as are GAA•TTC arrays (Krasilnikova and Mirkin, 2004) and CCG•CGG arrays (Pelletier *et al.*, 2003) in *S. cerevisiae* plasmid assays. Additionally, *in vitro* electron microscopy work with plasmid DNA has shown that TNR arrays may also regress to form a 4-way junction at the replication fork, suggesting a potential role of replication fork reversal in TNR instability (Fouche *et al.*, 2006).

Mirkin *et al* proposed a model by which a replication fork arrested at the site of a TNR array due to the presence of a hairpin-like structure on the lagging strand template, could produce either deletion or expansion instability (Mirkin, 2007). Such a stall could be resolved by Okazaki fragment skipping, leaving the region of the lagging strand unreplicated. If this region contained a TNR array folded into secondary structure, a subsequent round of replication would skip past this structure, producing a deletion in the TNR array (Figure 1.9). Furthermore, a pathway leading to expansion instability was proposed involving the stalled fork species. If stalling can cause a regression and reversal of the replication fork at TNR sequences, then reversal in such an intermediate would produce a region of single stranded nascent leading strand overhanging the newly synthesized lagging strand. This region would contain the TNR sequence present on the lagging strand template, and could thus be able to fold into a stable secondary structure. Restructuring of this reversed intermediate could then lead to the establishment of a functional replication fork in which the hairpin-like structure is present on the newly synthesized leading strand. Following a subsequent round of replication, this structure would produce an expansion in the TNR array (Figure 1.9).

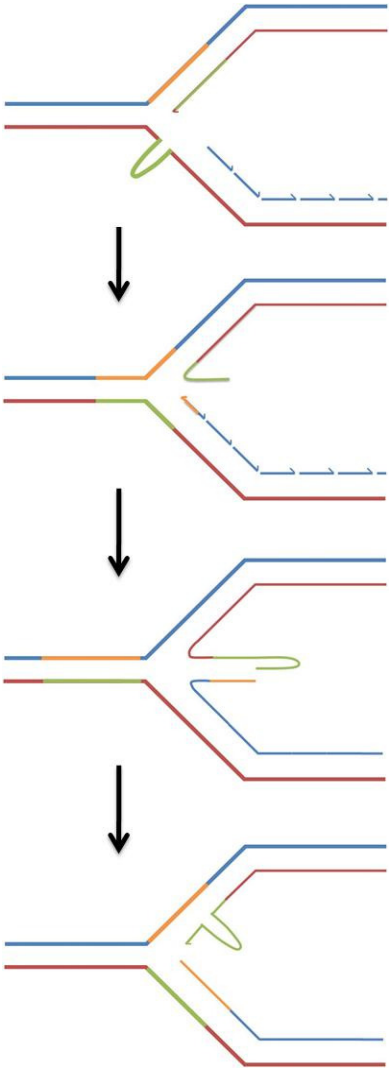

Figure 1.9. Replication Fork Reversal dependent expansion (adapted from (Mirkin, 2007)). A replication fork stalled at a TNR undergoes fork reversal producing single stranded region of the nascent leading strand. This repeat containing region can form a hairpin-like structure which, after fork restructuring, would produce an intermediate species from which a further round of replication would produce an expansion in the repeat array.

1.5.3 Recombination and TNR instability

Stalling of a replication fork could be stochastic in nature or could be induced by the presence of the TNR array itself. Evidence *in vitro* suggests that expandable

TNR sequences are able to stall replication fork progression (Kang *et al.*, 1995b) and *in vivo* in plasmids in the organisms *E.coli* (Samadashwily *et al.*, 1997) and *S. cerevisiae* (Pelletier *et al.*, 2003). Stalls to replication fork progression could be alleviated by a RFR mechanism (as above), or by a recombinational mechanism, either acting via the gap repair pathway, or if another replication fork had encountered a single stranded end on the stalled fork, a double strand break would be generated, requiring repair by the DSB HR repair pathway (Bichara *et al.*, 2006). Either of these HR repair pathways could generate instability in a TNR array at the repair site.

Plasmid studies in *E. coli* have established that TNR sequences can induce recombination (Napierala *et al.*, 2002; Pluciennik *et al.*, 2002). However, the studies have produced contradictory results for the effect of HR on instability in TNR arrays. Some studies have shown that a functional HR pathway destabilizes TNR array on plasmids (Hashem *et al.*, 2004; Jakupciak and Wells, 2000; Pluciennik *et al.*, 2002), while other studies have found the opposite – that HR stabilizes TNR carried on a plasmid, with instability increasing in HR deficient cells (Hebert *et al.*, 2004; Hebert and Wells, 2005; Sopher *et al.*, 2000). The use of plasmid based assays in these studies may be one problem associated with inconsistency of the results. These studies utilized different experimental methods, that in general either involved one multicopy plasmid in which TNR length was monitored in various mutant strains or a two plasmid system in which recombination between the two compatible plasmids was detected. Both of these systems have their own limitations: whether a mulitcopy plasmid reflects

accurately the mechanisms occurring in chromosomal DNA, instability occurring during transformation of the host cell with the plasmid DNA, and competition between the parental and recombinant plasmids in the two copy system (Bichara *et al.*, 2006; Bierne *et al.*, 1995; Hashem *et al.*, 2002).

Instability of TNR arrays could also occur after collapse of a replication fork, and subsequent repair. In yeast, long TNR arrays have been shown to be susceptible to breakage during the process of replication (Freudenreich *et al.*, 1998). A study into the role of HR repair of double strand breaks in *S. cerevisiae* found that an active HR system stabilized TNR arrays present on a yeast artificial chromosome (YAC) (Sundararajan *et al.*, 2010). However, other studies in *S. cerevisiae* strains mutated in the gene encoding recombination protein Rad52 showed no effect of the mutation on rates of expansion or deletion instability in otherwise wild type cells (Miret *et al.*, 1997, 1998). Studies in mice harbouring expanded CTG repeats, also saw no effect of mutation of the *Rad52* or *Rad54* recombination genes (Savouret *et al.*, 2003).

1.5.4 DNA repair and TNR instability

DNA lesions in a cell, either resulting from exogenous damage or from replication errors, can be repaired by a number of different processes. Many of the repair processes available to a cell involve some degree of DNA synthesis, meaning they are pathways in which TNR instability might occur.

Methyl-directed mismatch repair (MMR) is a pathway conserved from prokaryotes to eukaryotic life, and it is concerned with the repair of mismatches in DNA resulting in small (1-4bp) bulges in DNA (Bichara *et al.*, 2006). In *E. coli* the protein MutS recognises and binds a mismatch, then recruits the MutL protein, and activates the endonuclease protein MutH, which cleaves the DNA at a hemi-methylated GATC sequence. The resulting nick is extended by the UvrD helicase and exonuclease protein action, after which the gap can be refilled by polymerase III (Schofield and Hsieh, 2003).

Studies into the role of MMR in TNR instability in *E. coli* have produced some contradictory results. MMR was shown to increase instability of a $(CTG)_{180}$ TNR array present on the leading strand of plasmid DNA, as mutation in any of the *mutS, mutL* or *mutH* genes lead to less deletion instability compared to the wild type strain. However, the same study did not see the same effect when a CGG array was tested (Jaworski *et al.*, 1995). The opposite results were seen by another group studying TNR instability in a *mutS* mutant. In this case mutation of the *mutS* gene resulted in an increase in instability, particularly small changes in TNR length (Schumacher *et al.*, 1998).

Studies conducted in the eukaryotic *S. cerevisiae* have shown that a mutation in any of the genes encoding proteins involved in the MMR pathway in that organism, did not affect the frequency of large deletion or expansion events (Miret *et al.*, 1997, 1998). However, another study showed that a mutation of the gene encoding the MutS homologue MSH2, and another MMR protein PMS1, destabilized a CAG tract to specifically produce an increase in small (±1 TNR) changes in the tract length (Schweitzer and Livingston, 1997).

In higher eukaryotes repair mechanisms have been shown to play a role in the instability of TNR arrays, particularly of inherited expanded arrays in somatic non-dividing tissue, such as the tissues of the brain (Kovtun and McMurray, 2008). Instability of TNR in somatic cells is age-dependent, the number of repeats remains stable up to a certain stage in the lifespan of the animal, after which it can expand. In this way instability in somatic cells may affect the severity and onset of the disease (Kovtun and McMurray, 2001).

Work in transgenic mouse models carrying expanded TNR arrays have shown that mutation of the genes encoding MSH2 (homologous to MutS in *E. coli*) or MSH3 lead to a decrease in the number of expansion events in somatic cells (Kovtun and McMurray, 2001; Manley *et al.*, 1999). A complex of MSH2-MSH3 proteins, bind mismatched loops in DNA. This complex may bind the slipped strand structures formed by TNR arrays stabilising them enabling instability to occur (Kovtun and McMurray, 2001). However, a recent study investigating the role of human MutSβ (MSH2-MSH3) produced different

results. Using HeLa cell nuclear extracts *in vitro* it was shown that, in this system, MutSβ binding to CAG hairpins does not prevent the hairpin repair activities of the protein. The protein can still facilitate removal of the hairpin structure (Tian *et al.*, 2009). This study questions whether the results obtained from the transgenic mouse model applies to instability of CAG repeats in human cells.

Age-dependent instability of TNR arrays in somatic cells is also affected by the glycosylase OGG1 (8-oxo-guanine DNA glycosylase), which removes oxidized guanine residues from DNA (Kovtun and McMurray, 2008). In mouse models loss of the OGG1 protein leads to a reduction in the number of expansion events in somatic cells (Kovtun *et al.*, 2007). When OGG1 removes an oxidized guanine in a TNR array, a nick is generated in the DNA. During repair of this nick the DNA strand not being used as a template for synthesis will be displaced forming a flap, which should it contain the TNR array could fold into a secondary structure. Such a structure could be stabilized further by MSH2-MSH3 complex binding (Mirkin, 2007). FEN1 would normally be responsible for the removal of such flaps, however the formation of secondary structure could restrict FEN1 loading (Henricksen *et al.*, 2000; Spiro *et al.*, 1999).

Instability of TNR has been shown in both proliferative and non-proliferative cells. Work in transgenic mice has shown that in some tissues TNR expansions are age dependent. This seen in non-proliferative neuronal tissues indicating a DNA repair-based pathway of instability (Kovtun *et al.*, 2007;

Kovtun and McMurray, 2001). However, work investigating instability of CAG repeats in the germ cells of HD patients found that can expansions can arise in cells before the end of the first meiotic division. Expansions may be generated through replication-based mechanisms in these dividing cells (Yoon *et al.*, 2003). The use of *E. coli* as a model system to study instability of chromosomal TNR arrays is employed here. The bacterium is a rapidly dividing organism and as such will be used to investigate possible mechanisms of instability in proliferative tissue. In humans it may be that TNR instability is derived from a number of different mechanisms, with the proliferative nature of the cells at question being an important factor.

1.6 Work in this thesis

Work presented in this thesis aimed to investigate pathways *in vivo* of CTG•CAG TNR instability in *E. coli.* Most previous studies into TNR instability in the bacterium *E. coli* have been plasmid based assays, which have the limitations intrinsic to plasmid studies – effect of plasmid copy number, replicon size and instability arising during transformation of the host cell (Hashem *et al.*, 2002). A previous study in this laboratory involved an *in vivo* investigation into the instability of a chromosomal CTG•CAG TNR array, though this study focused mainly on deletion instability (Zahra *et al.*, 2007).

TNR arrays in yeast, close to the size of the threshold of 15 repeat units, have been shown to display a similar number of expansion and contraction

events (Rolfsmeier *et al.*, 2001). TNR arrays in bacteria, and longer TNR arrays in yeast cells show a bias towards deletion instability, whereas human cells with TNR greater than threshold length it is expansion instability that is responsible for TREDs (Kovtun and McMurray, 2008). However, a screen for yeast strains containing mutations which affected the ratio of expansions to deletions identified the gene *RTG2*. Yeast carrying a mutation in this gene undergo more expansion instability than deletion, much like the situation in higher eukaryotes (Bhattacharyya *et al.*, 2002). The aim of this thesis was to attempt to determine the mechanisms of both the deletion instability, and the less frequent, expansion instability pathways of chromosomal CTG•CAG TNR arrays in *E. coli.* This work was accomplished using a PCR based genetic assay developed previously in the laboratory.

The TNR arrays used in this study are $(CTG)_{95}$ and $(CAG)_{84}$, which are both beyond the threshold length of approximately 35 repeat units seen in human disease. This could be used to model the instability in such repeat sequences that are already beyond the threshold length. However, the relationship between TNR length and expansion instability was not investigated and whether the threshold length for the bacterium has been reached cannot be assumed. As expansions are rare in most of the strains studied it may be that in *E.coli* the threshold length is much higher than that in humans. The study describes how instability of an array is affected by particular cellular pathways and does not make conclusions about whether that instability is occurring in sub-threshold arrays, or those in which the threshold length has been exceeded.

The role of replication fork reversal on the instability of chromosomal CTG•CAG was investigated in Chapter Three. The helicase protein UvrD had previously been shown to be essential for RFR in a number of replication mutants in *E. coli* (Flores *et al.*, 2005). Here UvrD was shown to affect the deletion instability of CTG•CAG TNR arrays, though not through its role in RFR. No significant role of RFR was found on expansion instability of CTG•CAG TNR arrays.

In Chapter four, mutation in genes encoding proteins involved in gap repair and homologous recombination were tested for their effect on TNR instability. No effect of mutation of the genes *recF, recO* or *recA* on the frequency of deletion or expansion instability was found. However, in strains carrying a CTG TNR array on the leading strand, a mutation of these genes affected the size of deletion products produced, suggesting a role of the proteins encoded by these genes on the processing of intermediates of instability of TNR in this orientation.

Chapter five deals with the role of helicases on the instability of TNR arrays. The helicases UvrD and Rep were found to play a role in deletion instability of TNR arrays. RecG was shown to affect the size of deletion products produced, though not the frequency of deletion events in strains carrying a CTG array on the leading strand, suggesting a role of RecG in deletion intermediate processing. Rep, RecQ and DinG were found to play roles in expansion instability of strains carrying CAG TNR arrays on the leading strand. These results suggested a model in which helicases can play a role in

expansion instability of TNR arrays following arrest of the replication fork, with the particular helicase implicated being determined by the nature of the arrested fork.

The work presented here is a novel investigation into the instability of chromosomal CTG•CAG TNR arrays *in vivo* in *E. coli*. The findings of this study suggest a critical role for helicase action at arrested replication forks, resulting in instability of TNR arrays.

Chapter Two:

Materials and Methods

2.1 Materials

2.1.1 General Reagents

All chemicals and growth media were purchased from the following suppliers: Sigma, Fisher Scientific and Melford.

DNA polymerase enzymes used were *Taq* polymerase purchased from Roche and phusion polymerase produced by Finnzymes. Restriction enzymes were purchased from New England Biolabs.

2.1.2 Growth Media

2.1.2.1 Growth Media

Liquid growth media was autoclaved prior to use and stored at room temperature. Solid media was melted and autoclaved prior to use and stored at 55ºC.

Table 2.1 Growth Media

Media	Composition
LB Agar	1% Bacto-Tryptone, 0.5% Yeast Extract, 1% NaCl, 1.5% Bacto-Agar, pH adjusted to 7.2
L Broth	1% Bacto-tryptone, 0.5% Yeast Extract, 1%NaCl, pH adjusted to 7.2
LC Agar	1% Tryptone, 0.5% Yeast Extract, 0.5% NaCl, 1% Difco-Agar, pH adjusted to 7.2
LC Top Agar	1% Tryptone, 0.5% Yeast Extract, 0.5% NaCl, 0.7% Difco-Agar, pH adjusted to 7.2

2.1.2.2 Media Supplements

When supplements to media were used, the specified supplement was added to the growth media immediately prior to use.

Table 2. 1 Media Supplements

Antibiotic	Solvent	Stock Concentration	Final Concentration
Chloramphenicol (Cm)	100% Ethanol	50 ($mgml^{-1}$)	50 (μgml^{-1})
Tetracyclin (Tc)	50% Ethanol	15 ($mgml^{-1}$)	15 (μgml^{-1})
Kanamycin (Km)	Water	50 ($mgml^{-1}$)	50 (μgml^{-1})
Sucrose	Water	20%	5%

2.1.3 Buffers and Solutions

Table 2. 2 Buffers and solutions

Buffer	Composition
20x TAE	0.8M Tris, 0.4M Sodium Acetate, 0.02M EDTA, pH 8.2
Phage Buffer	7g Na_2HPO_4, 3g KH_2PO_4, 5g NaCl, 10ml $MgSO_4$ (0.1M), 10ml $CaCl_2$(0.1M), 1ml 1% Gelatine made up to 1Litre.
$CaCl_2$	Made fresh in Mili-Q water to 0.1M.
Sodium Citrate	Made up in Mili-Q water to 1M.

2.1.4 *Escherichia coli* Strains

Table 2. 3 *Escherichia coli* Strains

Strain	Background	Genotype	Source
MG1655		F⁻ lambda⁻ *ilvG*⁻ *rfb*-50 *rph*-1	(Blattner *et al.*, 1997)
XL1 Blue		recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac [F' proAB lacIqZΔM15 Tn10]	Stratagene
DL1786	MG1655	*LacZ* χ − *lacIq* ZeoR χ +	(Zahra *et al.*, 2007)
DL2009	DL1786	*LacZ*::$(CTG)_{95}$	(Zahra *et al.*, 2007)
DL2079	DL2009	*recA*::Cm	(Zahra *et al.*, 2007)
DL2099	DL2009	*recG*::Km (DL2009 x P1[DL1077])	John Blackwood
DL2384	DL2009	Δ*rep* (plasmid mediated deletion of *rep* gene using pDL2341)	John Blackwood
DL2629	DL2009	Δ*uvrD* (plasmid mediated deletion of *uvrD* gene using pDL2391)	John Blackwood
DL2639	DL1786	*lacZ*::$(CAG)_{84}$	(Zahra *et al.*, 2007)
DL2789	DL2639	Δ*uvrD* (plasmid mediated deletion of *uvrD* gene using pDL2391)	Rabaab Zahra
DL3455	DL2639	*recG*::Km (DL2009 x P1[DL1077])	This work
DL3672	DL2629	Δ*uvrD*Δ*recF* (plasmid mediated deletion of *recF* gene using pDL2748)	This work
DL3686	DL2009	*priC303*::Cm (DL2009 x P1[DL3674])	Ewa Okely
DL3687	DL2639	*priC303*::Cm (DL2639 x P1[DL3674])	Ewa Okely
DL3692	DL2639	Δ*rep* (plasmid mediated deletion of *rep* gene using pDL2341)	Ewa Okely

DL3742	DL2789	Δ*uvrD*Δ*recF* (plasmid mediated deletion of *recF* gene using pDL2748)	This Work
DL3891	DL2639	Δ*recF* (plasmid mediated deletion of *recF* gene using pDL2748)	Ewa Okely
DL3892	DL2009	Δ*recF* (plasmid mediated deletion of *recF* gene using pDL2748)	Ewa Okely
DL4148	DL2009	Δ*recO* (plasmid mediated deletion of *recO* gene using pDL2710)	Ewa Okely
DL4222	DL2009	Δ*dinG* (plasmid mediated deletion of *dinG* gene using pDL4187)	This Work
DL4223	DL2639	Δ*dinG* (plasmid mediated deletion of *dinG* gene using pDL4187)	This Work
DL4309	DL2009	Δ*recQ* (plasmid mediated deletion of *recQ* gene using pDL2765)	This Work
DL4310	DL2639	Δ*recO* (plasmid mediated deletion of *recO* gene using pDL2710)	This work
DL4337	DL2639	Δ*recQ* (plasmid mediated deletion of *recQ* gene using pDL2765)	This work
DL4341	DL2639	*recA*::Cm (DL2639 x P1[DL654])	This Work
DL4419	DL4222	Δ*dinG*Δ*rep* (plasmid mediated deletion of *rep* gene using pDL2341)	This Work
DL4420	DL4223	Δ*dinG*Δ*rep* (plasmid mediated deletion of *rep* gene using pDL2341)	This Work
DL4438	DL4309	Δ*recQ*Δ*rep* (plasmid mediated deletion of *rep* gene using pDL2341)	This Work
DL4487	DL4337	Δ*recQ*Δ*rep* (plasmid mediated deletion of *rep* gene using pDL2341)	This Work

2.1.5 Plasmids

Table 2. 4 Plasmids

Plasmid	Profile	Characteristics	Source
pTOF24	*repA*ts, *sacB*, *cat*, pSC101 ori	Used for SalI-PstI cloning	(Merlin *et al.*, 2002)
pDL2341	pTOF24-Δ*rep*	Used to integrate *rep* deletion	John Blackwood
pDL2391	pTOF24-Δ*uvrD*	Used to integrate *uvrD* deletion	John Blackwood
pDL2710	pTOF24-Δ*recO*	Used to integrate *recO* deletion	Ewa Okely
pDL2748	pTOF24-Δ*recF*	Used to integrate *recF* deletion	Ewa Okely
pDL2765	pTOF24-Δ*recQ*	Used to integrate *recQ* deletion	Ewa Okely
pDL4187	pTOF24-Δ*dinG*	Used to integrate *dinG* deletion	This Work

2.1.6 Oligonucleotides

Table 2.6 Oligonucleotides

The Fam-Ex-Test F primer was purchased from Metabion, all other oligonucleotides were purchased from MWG.

Name	Sequence (5'→3')	Description
Ex-Test F	TTATGCTTCCGGCTCGTATG	Used to amplify the TNR containing region of *lacZ*.
Ex-Test R	GGCGATTAAGTTGGGTAACG	
Fam-Ex-Test F	Fam-TTATGCTTCCGGCTCGTATG	
Rep 1F	AAAAACTGCAGTGAACGGTCCATTTTGGCTA	Used for the cross-over PCR

Rep 2F	CGTGTTATCACCGAGCGCAAAGTGGTCAGCGC	to produce the pTOF24-Δ*rep* plasmid (pDL2341) Primers F1 and R2 were used to confirm deletion of the gene.
Rep 1R	CACTTTGCGCTCGGTGATAACACGAGTTTTAC	
Rep 2R	AAAAAGTCGACATCAGCAAAATGCCGTACC	
UvrD 1F	AAAAACTGCAGTGACCTCGCTGATATAATCA	Used for the cross-over PCR to produce the pTOF24-Δ*uvrD* plasmid (pDL2391) Primers F1 and R2 were used to confirm deletion of the gene.
UvrD 2F	CGAGCCAGGCGGCGAGCTGCCGGAAGAGTGT	
UvrD 1R	CCGGCAGCTCGCCGCCCTGGCTCGTGCCCATC	
UvrD 2R	AAAAAGTCGACTCAGATACTGAAGATGGCGC	
RecF-KO-F1	AAAAACTGCAGCTCCAAACTGGTGGATGGTC	Used for the cross-over PCR to produce the pTOF24-Δ*recF* plasmid (pDL2748) Primers F1 and R2 were used to confirm deletion of the gene.
RecF-KO-F2	CCTCACCCGCTTGAAGATGTTTACCGTGGAAAAGG	
RecF-KO-R1	GTAAACATCTTCAAGCGGGTCAGGGACATTAC	
RecF-KO-R2	AAAAAGTCGACCGCCGGACACTTTATAGGAG	
RecO-KO-F1	AAAAACTGCAGCGTGCGTAAGCATCTACCTG	Used for the cross-over PCR to produce the pTOF24-Δ*recO* plasmid (pDL2710) Primers F1 and R2 were used to confirm deletion of the gene.
RecO-KO-F2	GATGGAAGGCTGACTGTTCCGGCAGTTTATGC	
RecO-KO-R1	TGCCGGAACAGTCAGCCTTCCATCGGAGTTAC	
RecO-KO-R2	AAAAAGTCGACCATCGCGCATTTTGTCAC	
DinG KO F1	AAAAACTGCAGTGTTCTGAGGTAGGCCAGGT	Used for the cross-over PCR

DinG KO F2	GCGTTTGTCGTACAGATGTCGCCCTTCTTCTC	to produce the pTOF24-Δ*dinG* plasmid (pDL4187) Primers F1 and R2 were used to confirm deletion of the gene.
DinG KO R1	GGGCGACATCTGTACGACAAACGCTTGCTGAC	
DinG KO R2	AAAAAGTCGACTTCGTAAAAGCCGAGCATCT	
RecQ-KO-F1	AAAAACTGCAGATGTGGTGGGTAATACTGACG	Used for the cross-over PCR to produce the pTOF24-Δ*recQ* plasmid (pDL2765) Primers F1 and R2 were used to confirm deletion of the gene.
RecQ-KO-F2	TGGAGTCCGGAGAACCGTTTATGGCGCTGATTC	
RecQ-KO-R1	CCATAAACGGTTCTCCGGACTCCAGATTCAAC	
RecQ-KO-R2	AAAAAGTCGACTTTTCAGTGCACCACGTAGC	

2.2 Methods

2.2.1 Bacterial methods

2.2.1.1 Growth of bacteria

Bacterial strains were streaked out from the -80ºC glycerol stock onto LB plates containing appropriate antibiotics to obtain single colonies. A colony was picked, used to inoculate 5ml of L broth, and grown overnight at the appropriate growing temperature with appropriate antibiotics.

2.2.1.2 Transformation of bacteria

Escherichia coli cultures were grown overnight and diluted 1:50 in LB broth and then grown for a further 2 hours at 37ºC so the culture was in the exponential growth phase. 1.5ml of culture was aliquoted into 2ml Eppendorf tubes and spun down in a microcentrifuge to pellet the cells. Following centrifugation the supernatant was discarded and 500µl of freshly prepared cold 0.1M $CaCl_2$ was added to resuspend the pellet, which was then left on ice for 30 minutes. The suspension was then centrifuged again and the supernatant discarded. The pellet was resuspended in 100µl of cold $CaCl_2$. Plasmid DNA was added and incubated on ice for 30 minutes, before heat shocking at 37ºC for 5 minutes. 400µl of LB broth was added and the cells incubated at an appropriate growth temperature for around 90 minutes of recovery to allow them to begin to express any antibiotic resistance genes. Cells were plated onto LB agar plates containing an appropriate antibiotic for selection of colonies transformed with the plasmid DNA and incubated overnight at the appropriate temperature.

2.2.1.3 P1 transduction

P1 is a bacteriophage virus that uses *E. coli* as its host. During the lytic growth phase of the phage's life cycle around 3% will mis-package *E. coli* DNA into the phage head instead of the viral DNA. This characteristic is exploited to transfer chromosomal material from one *E. coli* strain to another. When a second *E. coli* strain is infected by a mis-packaged phage, the region of genomic DNA homologous to the one carried by the phage can recombine with that from the

phage particle, creating a new strain. If a selectable marker, such as an antibiotic resistance gene, is in the region of DNA transduced from one strain to another, cells containing the new DNA can be isolated.

2.2.1.3.1 Making a P1 lysate

An overnight culture of cells was diluted 10-fold into 5ml of L broth supplemented with 2.5mM $CaCl_2$ and incubated at 37°C for 2 hours in a shaking incubator. 200µl of this culture was then mixed with 100µl of different dilutions of an old P1 lysate, and incubated for 30 minutes at 37°C. LC agar plates were made supplemented with 2.5mM $CaCl_2$. LC top agar supplemented with 2.5mM $CaCl_2$ was added to the cell and phage mixture. This was then poured onto the LC + 2.5mM $CaCl_2$ plates and left to set. The plates were incubated at 37°C overnight. The following morning the phage were collected by removing the top agar into 5ml of phage buffer. This solution was then vortexed with 100µl chloroform and left for 30 minutes at 4°C. Following centrifugation for 10 minutes, the supernatant was mixed with 100µl chloroform in a fresh bottle. The lysate was then stored in the dark at 4°C.

2.2.1.3.2 P1 transduction

An overnight culture of the *E. coli* strain due to receive the phage particle was grown overnight in 5ml of L broth + 2.5mM $CaCl_2$. 1ml aliquots of this culture were centrifuged for 1 minute and the supernatant discarded. The remaining

pellet of cells was resuspended in 100µl of fresh L broth + 2.5mM $CaCl_2$, and mixed with 100µl of differing dilutions of P1 lysate (or phage buffer only as a control). This was then incubated at 37°C in a shaking incubator for 30 minutes. 800µl of L broth supplemented with 2mM sodium citrate was then added and the mixture was then incubated for a further 1 hour at 37°C. 200µl of the cell culture was spread onto LB plates containing the appropriate selection antibiotic and incubated overnight at 37°C. Colonies grown on the selective plates were then streaked out and grown on selective plates twice more, to purify the bacteria away from the phage particles.

2.2.1.4 Plasmid mediated integration

Following a method established by Link and colleagues (Link *et al.*, 1997) precise alterations to the *E. coli* chromosome (deletions and insertions) were made using pTOF24 plasmid derivatives (Figure 2.1) (Merlin *et al.*, 2002). The plasmid contains a temperature sensitive replication protein (that allows growth at 30°C), a chloramphenicol resistance gene (*cat*), and the gene *sacB*, from the organism *Bacillus. subtilis*, which is lethal to *E. coli* in the presence of the sugar sucrose.

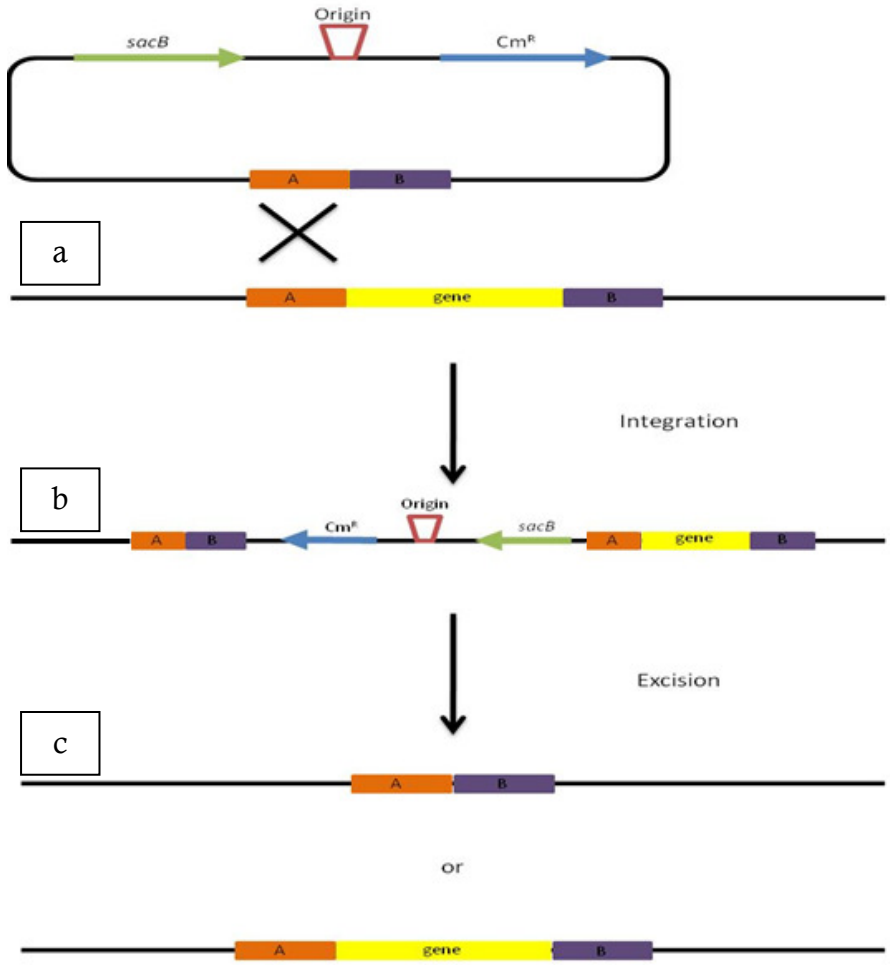

Figure 2.1: Plasmid mediated integration using a pTOF24 derivative plasmid to delete a gene. a. HR between the plasmid and the chromosome at homology arm A will integrate the plasmid. **b.** Incubating at 42°C on plates supplemented with chloramphenicol selects for cells in which integration has occurred. **c.** A second HR event can excise the plasmid DNA, selected for by growing on 5% sucrose plates. Cells in which HR occurs between the B homology arms results in a deletion of the gene, whereas when HR occurs between the A homology arms the gene would remain in the chromosome.

Products of a crossover PCR (section 2.2.2.7) in which two 400bp arms of chromosomal homology flanked by a Pst1 and a Sal1 restriction site, were digested and cloned into the pTOF24 plasmid vector (Figure 2.1a). The PCR product used determined the nature of the chromosomal alteration, two arms

designed to be homologous to the upstream and downstream regions of a gene with nothing in between will produce a deletion of that gene by this method, conversely PCR products with two arms of chromosomal homology with DNA sequence added between them will insert that sequence into the genomic DNA at the site of homology. Receiving *E. coli* strains were then transformed with the newly constructed pTOF24 plasmid.

Plasmids containing homology to chromosomal regions of DNA can integrate into the bacterial chromosome by homologous recombination. The transformed strains were streaked onto LB plates supplemented with chloramphenicol for selection, and grown overnight at 42 ºC. This temperature is non-permissive for the replication of the plasmid, so selectively allows growth of cells in which the plasmid has integrated into the chromosome (Figure 2.1b). Colonies that had integrated the plasmid were picked, streaked onto LB plates supplemented with chloramphenicol and grown again overnight at 42 ºC to purify them.

The integrated plasmid can also be excised by homologous recombination and selected for by using the sugar sucrose which confers lethality to *E. coli* cells expressing the *sacB* gene. Liquid cultures in L broth, of colonies with pTOF24 plasmid integrated into their chromosomes, were inoculated and grown overnight at 30 ºC with no selection. 100µl of 10^{-5} dilutions of these overnight cultures were then plated onto LB plates supplemented with 5% sucrose. The resulting growing colonies will have excised the plasmid, which was confirmed by testing for death on LB plates supplemented with chloramphenicol. One of

two possible excision events could have taken place (Figure 2.1c), one in which the original DNA was retained or one in which the newly modified DNA was retained. This was tested for by PCR screening of the region.

2.2.1.5 Phenotypic UV test

RecA$^-$ strains are sensitive to damage by UV irradiation. Strains in which a *recA* mutation had been introduced by P1 transduction were tested for the mutation using a UV phenotypic test. An overnight culture of the strain was serially diluted to 10^{-6} and 10µl of each dilution was spotted onto two LB plates in parallel, with controls known to be *recA*$^+$ and *recA*$^-$. One plate was then exposed to $1Jm^{-2}$ UV irradiation using the UV Stratalinker™ 1800 (Stratagene). Both plates were then incubated overnight in the dark at 37°C, and the growth of the potential mutant strain was compared to the controls.

2.2.1.6 Storage of bacteria

Bacterial strains were stored in a strain collection at -80ºC. A strain was grown overnight in 5ml L broth at the appropriate growth temperature, then 700µl of this overnight culture was mixed with 700µl of 80% glycerol in a 1.5ml Eppendorf tube, before being added to the strain collection.

2.2.2 DNA methods

2.2.2.1 Boiled cell DNA isolation

The desired strain was streaked onto an LB agar plate to obtain single colonies. A colony was picked and suspended in 50µl of sterile water. The cell suspension was then held at 99ºC for 10 minutes in a PCR machine to lyse the cells releasing the DNA into the solution. This was then centrifuged for 2 minutes to separate the cell debris. 2µl of the supernatant was then used as template DNA in a 25µl PCR reaction.

2.2.2.2 Plasmid DNA isolation

Plasmid DNA was isolated from a 5ml overnight culture of the desired strain, grown to stationary phase, using the QIAprep Spin Miniprep Kit (Qiagen). Manufacturer's instructions were followed. The plasmid DNA was eluted using 30µl of elution buffer, and stored at -20ºC until required.

2.2.2.3 Genomic DNA isolation

Genomic DNA was isolated from a 5ml overnight culture of the desired strain, grown to stationary phase, using the Wizard Genomic DNA purification kit (Promega). Manufacturer's instructions were followed. The genomic DNA was stored at -20ºC until required.

2.2.2.4 Restriction digestion

Restriction digests were carried out using enzymes from New England Biolabs following manufacturer's instructions for the particular enzyme. The required reaction conditions varied for different enzymes but in general reactions contained DNA, the appropriate restriction buffer and BSA where needed, and 2-5 units of enzyme. Digests were incubated at the appropriate temperature for the enzyme for 1-3 hours. Digestion products could then be analysed on 1% agarose gels (section 2.2.2.9) to check fragments produced.

2.2.2.5 DNA ligation

DNA ligations were carried out using the Quick Ligation kit (New England Biolabs), following the manufacturer's instructions.

2.2.2.6 PCR

Polymerase chain reaction (PCR) was used to amplify specific regions of DNA from either plasmid or genomic DNA. In general a PCR reaction of 25µl was setup as follows:

2µl	Template DNA
1µl	Forward primer (5pmol μl^{-1})
1µl	Reverse primer (5pmol μl^{-1})
2.5µl	dNTP mixture (2mM)
2.5µl	10x PCR buffer
0.2µl	DNA polymerase
15.8µl	Sterile water

Taq polymerase (Roche) was used in most cases. For PCR reactions requiring a proofreading activity, the enzyme *Phusion* polymerase (Finnzymes) was used, following manufacturer's instructions. PCR reactions were carried out in a PCR machine using a program based on the following, where T_m is the melting temperature of the primer:

95ºC	5 minutes	
95 ºC	30 seconds	x30 cycles
T_m - 5 ºC	15 seconds	
72 ºC	1minute/kb DNA	
72 ºC	10 minutes	
8 ºC	∞	

2.2.2.7 Crossover PCR

Crossover PCR is a technique used to produce DNA fragments with chromosomal homology arms used in construction of pTOF24 vectors needed for plasmid mediated integration (Section 2.2.1.4), (Figure 2.1), (Ho *et al.*, 1989). Two pairs of primers are used, two forward and two reverse (Figure 2.2a). The two external primers (forward 1 and reverse 2) contain restriction sites, while the two internal primers (forward 2 and reverse 1) contain a 24bp region of homology to each other.

The first two PCR reactions take place separately, one using the forward 1 and reverse 1 primers, the other reaction using forward 2 and reverse 2 primers (Figure 2.2b). The products were then purified and mixed together as the DNA template for a third amplification reaction using the forward 1 and reverse 2 primers, producing one fragment – a crossover product of the two reactions. The restriction sites could then be used to enable this fragment to be cloned into the pTOF24 vector.

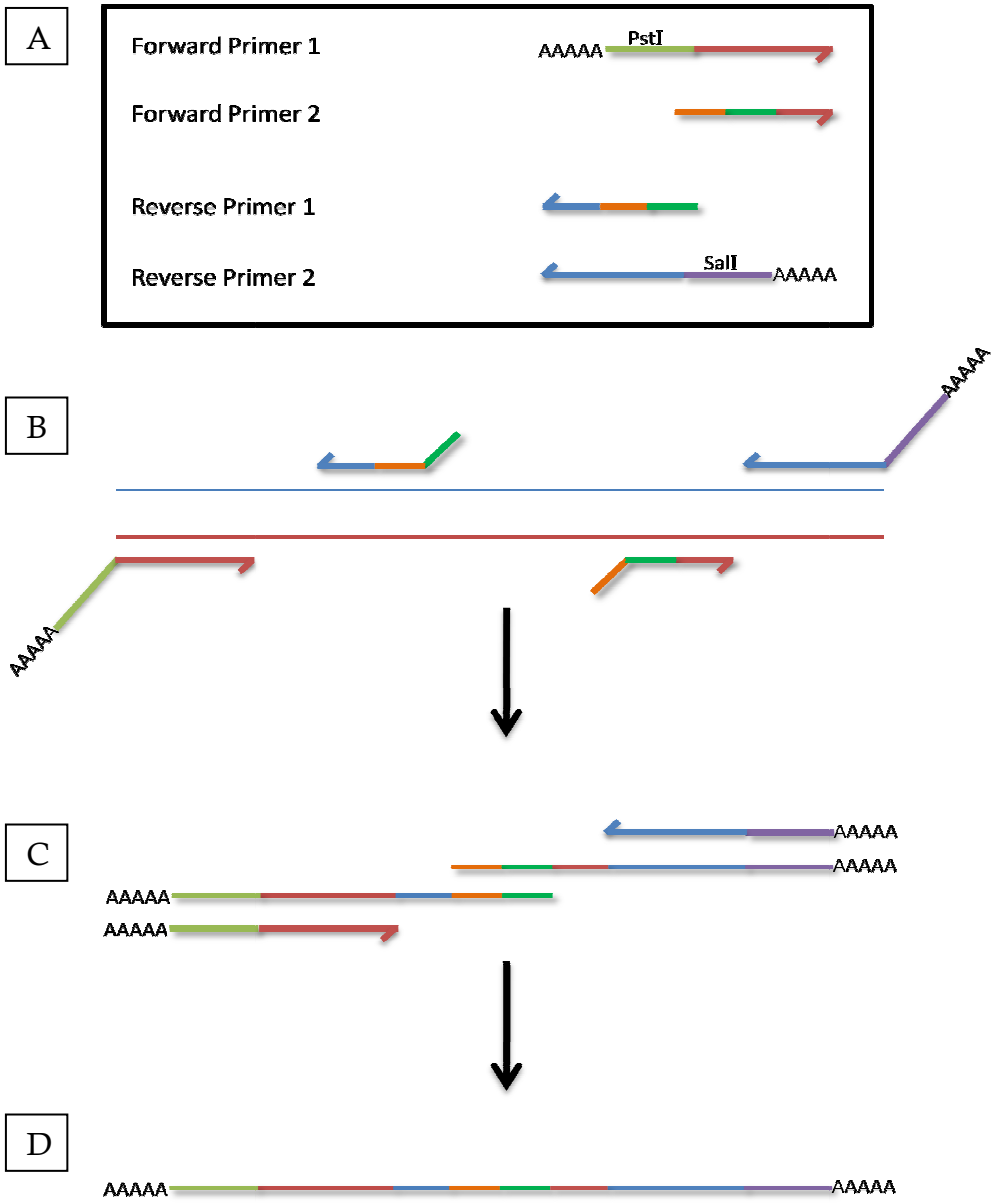

Figure 2.2: Crossover PCR reaction. A. Primers used in crossover PCR reaction. **B.** Two separate PCR reactions conducted. **C.** Third PCR reaction using the outside primers on the products of the PCR reactions from **B**. **D.** Crossover PCR product.

2.2.2.8 DNA sequencing

DNA to be sequenced was first purified using a QIAquick PCR purification kit (Qiagen). The sequencing reaction was then carried out using the BigDye

Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) following the manufacturer's instructions. Samples were sent to The Gene Pool, University of Edinburgh where they were run on an ABI PRISM 3100-Avant Genetic Analyzer.

2.2.2.9 Gel electrophoresis

DNA was analysed for size on agarose gels. 1% agarose gels were made by melting 5g of agarose in 500ml of 1x TAE buffer in the microwave. Safe View (NBS Biologicals) was added to the melted agarose after it had cooled slightly, to visualize the DNA once the gel had been run. 5µl of safe view was added per 100ml of agarose. Agarose was then poured into gel trays and allowed to set. 5µl samples of DNA to be run on the gels was added to 2µl of 60% glycerol and loaded into the wells of the agarose gel. The gel was run submerged in 1x TAE buffer with a potential difference applied across it. DNA was visualized using a UV trans-illuminator. DNA markers purchased from New England Biolabs were used to quantify the size and quantity of DNA in bands visualized.

2.2.3 Genetic Assay

CTG•CAG TNR arrays were inserted into the bacterial chromosome in two different orientations relative to the origin of replication (Zahra *et al.*, 2007). When an array was inserted in such an orientation, that CTG repeats were

present on the leading strand of the duplex DNA, and CAG repeats were therefore present on the lagging strand, that was termed the CTG orientation. When CAG repeats were present on the leading strand, and CTG on the lagging strand, this was termed the CAG orientation. Analysis of the instability of TNR arrays in the *E. coli* chromosome was conducted using a genetic assay (Figure 2.3) designed in the laboratory by Rabaab Zahra.

2.2.3.1 Instability assay

Strains on which the instability assay was to be performed were streaked out from the -80ºC stock onto LB plates to produce single colonies. 60 single colonies were then selected for each strain and each colony was inoculated into 5ml of L broth and grown overnight in a shaking incubator at 37ºC. The instability of the TNR array at this stage of growth was assessed. A 10^{-6} dilution of the overnight cultures was produced in L broth, and 100µl of this dilution was plated onto LB agar plates. Plates were incubated overnight in a 37ºC incubator to allow growth of the colonies. 8 colonies were then selected from each plate and PCR carried out to determine the length of the TNR array in each.

The PCR reaction using Ex-Test primers was designed to amplify the region of the *lacZ* gene in which the TNR array was inserted. For each PCR reaction, the reverse primer Ex-Test R and a modified version of the forward primer Ex-Test F which was labelled with a fluorescent 5'-FAM (Fam-Ex-Test F), were used.

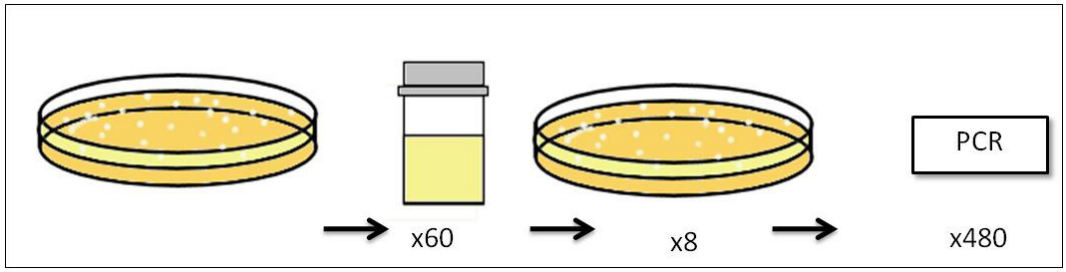

Figure 2.3: Representation of the TNR instability assay. 60 overnight cultures were setup for each strain tested and grown at 37°C. 10^{-6} dilutions were made and plated onto LB plates. 8 colonies from each plate were then tested for repeat length by PCR.

2.2.3.2 GeneMapper analysis

Products of the PCR reaction were diluted 1:25 in sterile water. 1ml of HiDi™ Formamide was then mixed with 5µl of GeneScan™-1200 LIZ™ size standard, and 9µl of this solution was mixed with 1µl of the diluted PCR products in bar-coded 96 well plates. Plates were sent to The Gene Pool, The University of Edinburgh, where they were run on the ABI 3730 Genetic Analyzer, which separates DNA fragments through capillary electrophoresis.

Data produced were analysed using the software GeneMapper v3.7. The data relates the time taken for the DNA fragments in the PCR products to pass a laser (x-axis). The software compares this time to that of the size standard and converts time to size, allowing sizing of the DNA fragments. Using this method the size of TNR arrays could be determined to ±5bp. Figure 2.4 shows examples of GeneMapper results obtained by this method.

Figure 2.4a shows a parental peak surrounded by a number of 'stutter' peaks, a typical output of the GeneMapper program. These 'stutter' peaks are *in vitro* artefacts from the PCR reaction and not real instability products. The size of the TNR array can be calculated from the fragment size detected by subtracting 148bp (the size of the region outside the repeat array amplified by the Ex-Test primers) and dividing by 3, to give the number of TNR units. Colonies producing results such as displayed in Figure 2.4a were considered to have not undergone an instability event, as the parental length of the TNR array was detected. The graph in Figure 2.4b shows the results of a colony in which two DNA fragment lengths were detected. Such colonies represent mixed populations, and to avoid counting instability events that have occurred on the plate only they were considered to be parental. Colonies in which a deletion event had occurred in the overnight culture, reducing the size of the TNR array produced GeneMapper outputs as in Figure 2.4c, while those in which expansion events occurred in the overnight culture produced outputs as in Figure 2.4d.

On rare occurrences when two colonies from the same overnight culture produced the same instability length, only one was counted as they were considered to be sister clones. Plates in which four or more of the eight colonies tested were sister clones were not included in the analysis. To ensure that the colony from which the overnight culture was produced had the expected TNR array size at least one colony of parental length had to be present in the eight tested, or that culture was not included in the analysis.

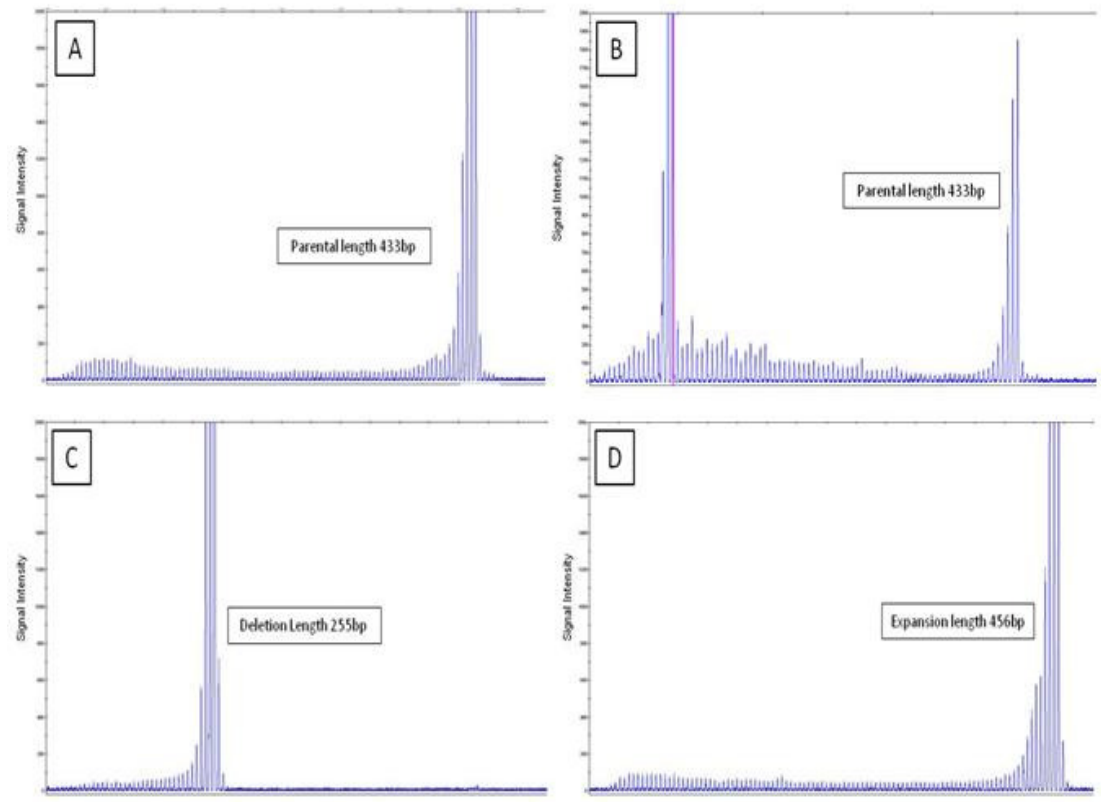

Figure 2.4: Example of GeneMapper output. A. The peak represents the parental length of $(CTG)_{95}$ TNR array, 433bp ((3bp x 95) + 148bp = 433bp). The region amplified outside the repeat array is 148bp. **B.** Outputs displaying both parental and deletion peaks were considered to be result of deletion events occurring during the growth of the colony on the agar plate. **C.** The peak represents a deletion of the TNR array to $(CTG)_{36}$. **D.** The peak represents an expansion of the TNR array to $(CTG)_{103}$.

2.2.4 Statistical methods

Graphs and statistics were produced using the software Microsoft Excel 2007 and Minitab 15, respectively.

The proportion of instability events resulting in a deletion of the TNR array, and the proportion of instability events resulting in an expansion of the

array, were calculated for each of 60 overnight cultures produced from a particular *E. coli* strain being investigated. The proportion of instability events calculated for each culture was treated as sample data, the mean of the samples was then taken to determine the population mean for the strain. The mean proportion of deletion events calculated over the population for a particular strain will hereafter be termed the deletion instability proportion. The mean proportion of expansion events calculated over the population for a particular strain will hereafter be termed the expansion instability proportion. Standard deviation within the population was then calculated, and from that the standard error of the mean found using the calculation:

$$\text{Standard Error of Mean} = \frac{\text{Standard deviation}}{\sqrt{n}}$$

Population means were plotted for each strain investigated and error bars drawn showing the standard error of the mean for each.

The instability assay data was considered as a binary response, whether an instability event had occurred or not, and was scored accordingly: 1 = instability event, 0 = no event. A logistic regression model was applied to the data to compare the instability in mutant strains to the corresponding wild type strain for that orientation of TNR array. Logistic regression is used in situations in which the dependent variable produces binary responses - in this case whether an instability event had occurred (=1) or not (=0). Linear regression cannot be utilized in such cases as predicted values of >1 and <0 will be generated. Linear regression models assume homoscedasticity of the data, meaning the variance of

Y is constant across all values of X, this is not the case for binary response data. The logistic curve instead relates the independent variable X to the rolling mean of the dependent variable. This produced a probability (P-value) as to whether a mutant strain had a different instability proportion to the wild type.

Median and inter-quartile expansion lengths were calculated for each strain, and plotted. The non-parametric Kruskal-Wallis test applied to test for differences in expansion sizes produced between strains. The Kruskal-Wallis test is a non-parametric equivalent of the parametric test ANOVA (analysis of variants) and is used to test whether population medians between different groups are equivalent or not. Unlike ANOVA analysis it does not assume that the populations are normally distributed, and does not require a large sample size.

The size of deletion instability events detected was also analysed. Deletion events, both those considered to have occurred during growth of the culture over night and those produced during growth of the culture on the LB plate, were plotted as size of the deletion product, expressed as a percentage size of the parental repeat array, against the number of such size products detected. Trend-lines were then fitted to the data. A quartic trend-line was used to model the data, this was chosen as it produced the best value of correlation coefficient for all the strains investigated, where the value of r^2 was as close to $r^2=1$ as possible. The average (median) size of deletion product produced for each strain was also determined.

Chapter Three:

Replication Fork Reversal And Trinucleotide Repeat Instability

3.1 Introduction

This chapter describes studies performed to determine the effect of replication fork reversal on CTG·CAG repeat instability in the *E. coli* chromosome. As discussed in chapter one Mirkin and collaborators proposed a model suggesting that expansion instability of TNR is induced by a fork reversal mechanism (Figure 1.9) (Mirkin, 2007). Replication fork reversal is a process by which an arrested replication fork can be restructured so it is capable of re-priming replication, by annealing of the two nascent strands, followed by degradation or recombination (Seigneur *et al.*, 1998). The aim of the work presented in this chapter was to test this hypothesized model, to find the effect that fork reversal might have on expansion and deletion instability of CTG•CAG TNR *in vivo*.

3.1.1 UvrD Helicase

UvrD, also known as helicase II, is a superfamily 1 helicase that acts in many different processes in the *E. coli* cell. It plays roles in the nucleotide excision repair and the methyl-directed mismatch repair pathways (Matson and Kaiser-Rogers, 1990) where it unwinds the lesion-containing strand to allow repair to take place. It also has a role in clearing RecA protein from single stranded DNA, both *in vitro* and *in vivo* (Figure 1.5) (Centore *et al.*, 2009). This affects the process of homologous recombination (HR). Indeed a *uvrD* mutant exhibits a hyper-recombination phenotype, while cells in which the UvrD protein is over expressed have a decreased rate of HR (Centore *et al.*, 2009; Zieg *et al.*, 1978).

3.1.2 RecF Protein

RecF protein is part of the RecFOR complex, and is most commonly associated with repair of gapped DNA structures in the RecF pathway of repair (Figure 1.7). The role of RecF protein is to allow binding of RecA protein to single stranded DNA coated in SSB protein (Morimatsu and Kowalczykowski, 2003).

3.1.3 UvrD, RecF and RFR

Studies using DnaE and DnaN temperature sensitive mutants to induce replication fork arrest have established that in such stalls the helicase UvrD is essential for the replication fork reversal process in otherwise wild type cells (Flores *et al.*, 2004). In this instance, the UvrD protein is believed to remove the RecA filament from single stranded DNA, thereby allowing replication fork reversal to take place (Flores *et al.*, 2005). RecA protein is loaded onto single stranded DNA at arrested forks by the RecFOR proteins (Morimatsu and Kowalczykowski, 2003). *dnaE*ts or *dnaN*ts cells lacking both UvrD and one of the RecFOR proteins are viable, as the RecA protein cannot be loaded onto the single stranded DNA of these arrested forks, so fork reversal and eventual replication resumption do no longer require the presence of UvrD (Flores *et al.*, 2005).

The role of UvrD in *E. coli* is shared by the closely related protein Srs2 found in the eukaryote *Saccharomyces cerevisiae*. *In vitro* protein is able to disrupt

the filaments of Rad51 protein, the homologue of the *E. coli* RecA, that form on single stranded DNA (Veaute *et al.*, 2003).

The roles of the UvrD and RecF proteins in this RFR pathway were utilized here to investigate the effect of this pathway on the instability of CTG•CAG TNR arrays in the *E. coli* chromosome.

I show here that while the protein UvrD does have an effect on the deletion instability of TNR arrays, it does not significantly affect the expansion instability pathway.

3.2 Results

3.2.1 Deletion instability in RFR mutants

3.2.1.1 Deletion proportion

The average proportion of deletion events (deletion instability proportion) was calculated from 60 independent assays for each *E. coli* strain investigated (Figure 3.1). In the mutants studied deletion instability was dependent on the orientation of the repeat array relative to the origin of replication, consistent with previous observations (Kang *et al.*, 1995a). This finding supports a model in which deletion instability occurs during DNA replication. A TNR array present in the transiently single stranded region of the lagging strand template may form secondary structure (Figure 1.8). The more thermodynamically stable structure forming sequence (CTG) is present in this position in the so-called 'CAG

orientation'. The increased thermodynamic stability of the CTG hairpin sequence would enable more structures to survive into a second round of replication, leading to a deletion in the repeat array. The less thermodynamically stable structure forming CAG sequence may not form hairpins at all, but potentially forms looped out DNA which has less chance of surviving into a second round of replication, resulting in fewer deletion events.

As shown in Figure 3.1, deletion of the *uvrD* gene increased the deletion proportion by approximately 2-fold in the CTG leading strand orientation and by approximately 3-fold in the CAG leading strand orientation. Logistic regression analysis of these data (Table 3.1) supports this observation, as statistical significance was found in the Δ*uvrD* mutants containing the TNR in either orientations ($p= 0.048$ and $p<0.001$ for $(CTG)_{95}$ and $(CAG)_{84}$ respectively).

A slight stabilising effect of a Δ*recF* mutation in a Δ*uvrD* mutant background could be seen in a strain containing the $(CAG)_{84}$ array, as the deletion proportion was reduced in the Δ*uvrD*Δ*recF* mutant compared to the Δ*uvrD* background strain. In the strain containing the $(CTG)_{95}$ array this difference was less pronounced. In this orientation much smaller numbers of deletion events were being encountered, so an effect may be unnoticeable.

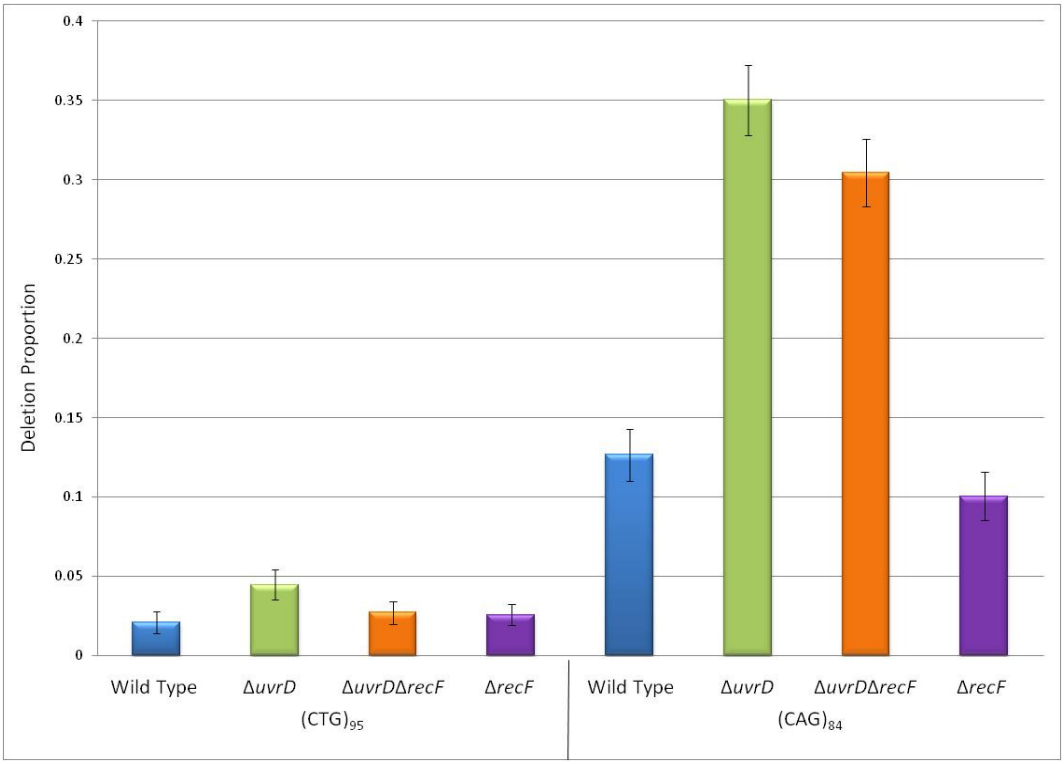

Figure 3.1: Deletion proportion of CTG•CAG arrays in *E. coli* strains containing a mutation in genes involved in the replication fork reversal pathway. Each bar represents the proportion of deletion instability events calculated from 60 independent assays. Error bars show the standard error of the mean calculated for each strain.

Table 3.1 Logistic regression analysis of deletion instability proportions of RFR mutant strains. p≤ 0.05 considered statistically significant.

Orientation	Strain	P-Value
$(CTG)_{95}$	Δ*uvrD*	0.048
	Δ*uvrD*Δ*recF*	0.528
	Δ*recF*	0.659
$(CAG)_{84}$	Δ*uvrD*	<0.001
	Δ*uvrD*Δ*recF*	<0.001
	Δ*recF*	0.183

3.2.1.2 Deletion size

All deletion lengths detected, both those considered to have occurred during the growth of the overnight culture and those produced from colony growth on the plate, were converted to a percentage size of the parental TNR array length. The frequency of each new TNR length produced was counted and plotted (Figure 3.2 and Figure 3.3 $(CTG)_{95}$ and $(CAG)_{84}$ respectively). Data were plotted for each of the strains individually (panels A-D in both figures) along with the median value deletion size. Trend lines were drawn for each strain and plotted together (panel E both figures).

Wild type cells carrying TNR in the CTG orientation had an average (median) of 35% of the parental array remaining after the deletion event (Figure 3.2). Mutation of either the *uvrD* or *recF* genes resulted in smaller deletion sizes (41% and 42% respectively), while a double mutant strain was even more affected (45%). Similarly, for strains carrying a TNR array in the CAG orientation (Figure 3.3) a mutation in *uvrD* affected the median deletion size (38% compared to the wild type 30%). However, in strains carrying TNR arrays in this orientation the effect of Δ*recF* is less pronounced (33%).

The deletion distribution of strains carrying TNR array in the CAG orientation (Figure 3.3) showed that while different median values were found between mutant strains, the deletion distributions were very similar. Trend lines fitted to the data all displayed a skew towards big deletion events. However, strains carrying TNR arrays in the CTG orientation (Figure 3.2) displayed

deletion distributions which were similar only for the wild type and Δ*uvrD* strains (though more events were detected in the Δ*uvrD* mutant). The distributions of strains containing a Δ*recF* mutation appeared different. These mutants displayed a more even distribution, less biased towards big deletion events, though this could be an artefact from the fact that in this orientation of repeat array there was a lower frequency of deletion events, resulting in a trend line which fits the data less well than in the CAG leading strand orientation.

The median values for deletion instability of the strains investigated here cannot be compared statistically to test for significance. The data are not normally distributed, and so are not suitable for parametric analyses. Non-parametric methods are also not appropriate as such analyses assume low values of n ($n<100$) and some of the values of n in these data exceed 100. While statistical significance cannot be established, the data on median deletion size and the deletion distributions produced suggest that strains carrying a CTG TNR array on the leading strand template with a mutation in the *recF* gene produce smaller deletion products than strains in this orientation that are *recF*$^{+}$. This result is supported by similar results in the strains with a mutation in *recA* and *recO* (Figure 4.3).

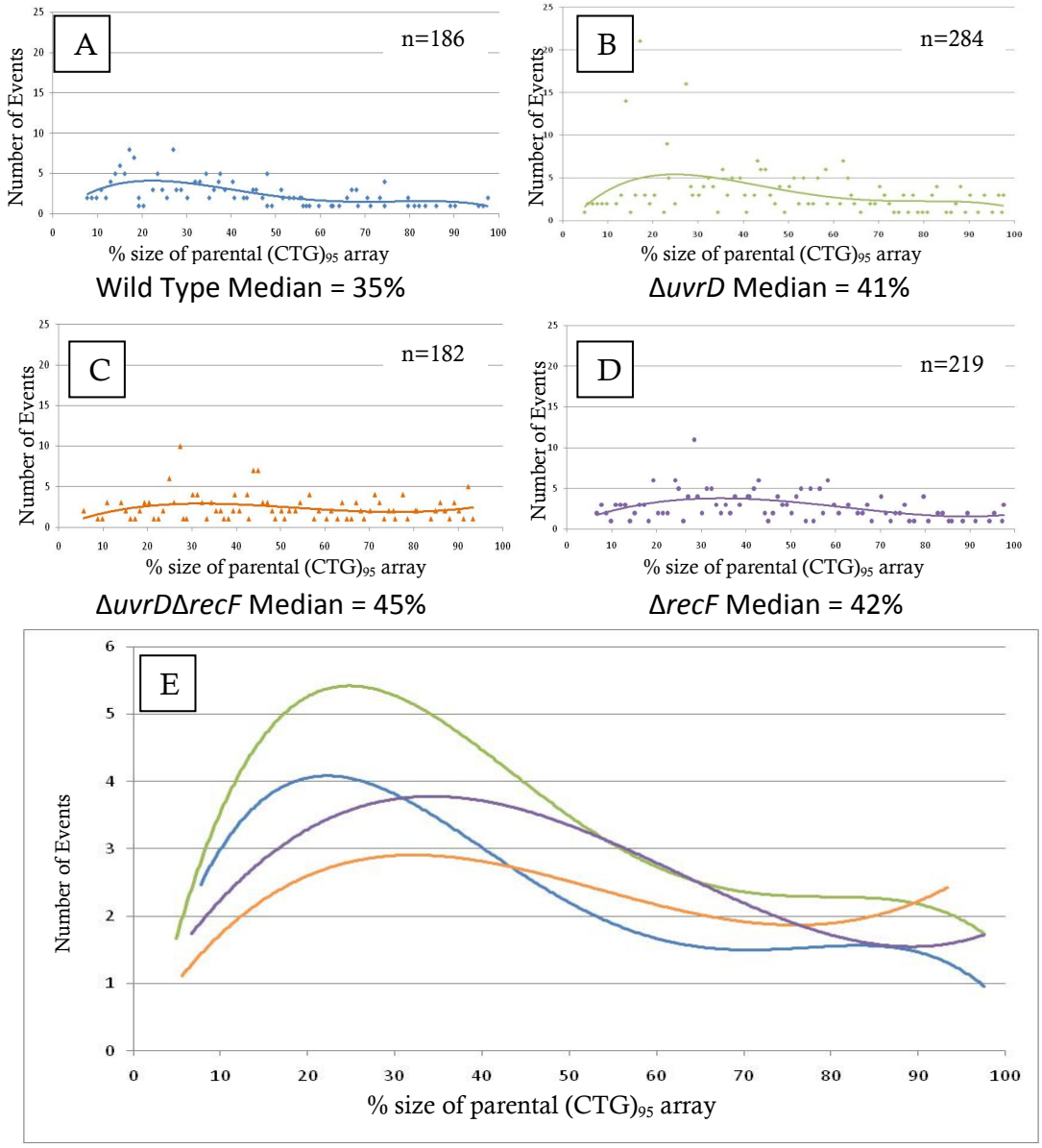

Figure 3.2: Distribution of deletion sizes of *E. coli* strains carrying the $(CTG)_{95}$ repeat array on the leading strand. Deletions are presented as a percentage of the parental 95 repeat array plotted against the frequency of the observed deletion size for strains: Wild Type (**A**), Δ*uvrD* (**B**), Δ*uvrD*Δ*recF* (**C**) and Δ*recF* (**D**). Trend lines for each distribution were plotted together (**E**).

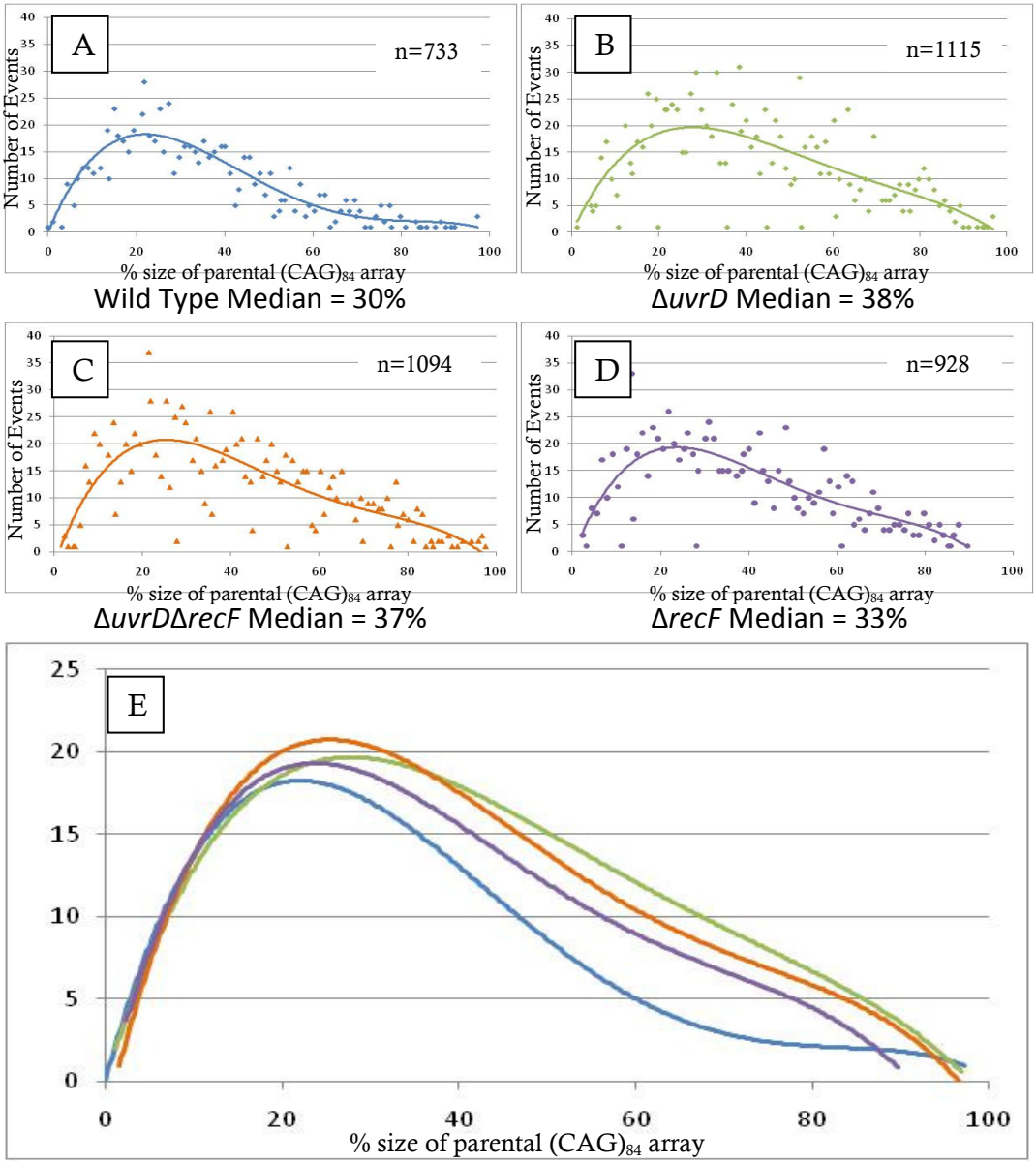

Figure 3.3: Distribution of deletion sizes of *E. coli* strains carrying the $(CAG)_{84}$ repeat array on the leading strand. Deletions are presented as a percentage of the parental 84 repeat array plotted against the frequency of the observed deletion size for strains: Wild Type (**A**), Δ*uvrD* (**B**), Δ*uvrD*Δ*recF* (**C**) and Δ*recF* (**D**). Trend lines for each distribution were plotted together (**E**).

3.2.2 Expansion instability in RFR mutants

3.2.2.1 Expansion Proportion

The average expansion instability proportion was calculated from 60 independent assays for each strain investigated (Figure 3.4). Unlike the deletion instability, expansion instability in these strains appears to be orientation independent. Notably, the strain containing the $(CAG)_{84}$ repeats on the leading strand and mutated in the *uvrD* gene displayed an elevated level of expansion instability. However, when a logistic regression model was applied to these results (Table 3.2), no significant difference was revealed between the wild type strains and any of the mutant strains in either orientation of the repeat array. The $(CAG)_{84}$ Δ*uvrD* strain was not significantly different to the wild type strain carrying the TNR in the same orientation ($p=0.213$) and the instability of that strain was almost the same as the instability of the $(CTG)_{95}$ Δ*uvrD* strain. Therefore the data show that in either orientation there is no statistically significant effect of *uvrD* or *recF* deletion on CTG•CAG expansion instability.

The assay utilized in this study was originally designed to detect deletion instability in the *E. coli* chromosome. Deletion instability occurs at a much higher frequency than expansion instability, so the assay itself may not accurately capture the differences of expansion instability within the strains.

In this assay, colonies in which parental and instability repeat lengths were both detected were considered to have remained parental during growth of the overnight culture, with the detected instability having occurred while those

colonies grew on the agar plate. For this reason such colonies were previously counted as parental length TNR arrays. However, such instability events may help to provide a clearer picture of the propensity of a particular strain to expand, so could be included in the analysis. Therefore expansion instability proportions were recalculated from each of the 60 independent assays for each strain, this time including expansions previously considered to have occurred on the plate (Figure 3.5). Colonies in which a deletion event had taken place in the overnight culture (containing no parental length) were not included in the denominator when calculating expansion proportions. Logistic regression analysis was also carried out on these results (Table 3.3).

The outcome of this second analysis confirmed the results obtained in the first. Though the data in Fig 3.5 would suggest an orientation dependence in expansion instability, this was an artefact due to the exclusion of colonies in which deletion events took place in the overnight culture. The orientation dependence of deletion instability affected the results in this analysis. The $(CAG)_{84}$ repeat showed a higher level of deletion instability than the $(CTG)_{95}$ repeat which was reflected accordingly on the level of expansion instability. Logistic regression analysis of the data seemed to suggest a potential effect of *uvrD* deletion on expansion instability in the $(CAG)_{84}$ orientation strains. However, this observation was again the reflection of significantly more deletion events occurring in this strain, which artificially increased the proportion of expansion instability events.

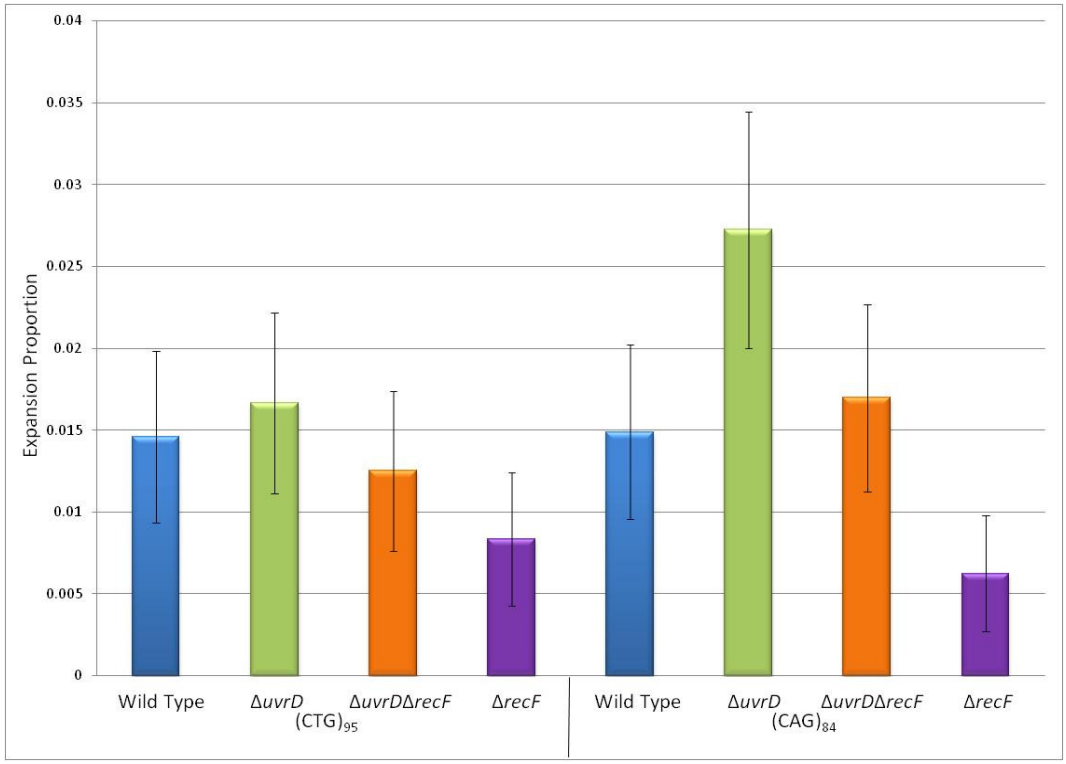

Figure 3.4: Expansion instability proportion of CTG•CAG arrays in *E. coli* strains containing a mutation in genes involved in the replication fork reversal pathway. Each bar represents the proportion of expansion instability events calculated from 60 independent assays. Error bars show the standard error of the mean calculated for each strain.

Table 3.2: Logistic regression analysis of expansion instability proportions of RFR mutant strains. p≤ 0.05 considered statistically significant.

Orientation	Strain	P-Value
$(CTG)_{95}$	Δ*uvrD*	0.785
	Δ*uvrD*Δ*recF*	0.780
	Δ*recF*	0.373
$(CAG)_{84}$	Δ*uvrD*	0.213
	Δ*uvrD*Δ*recF*	0.763
	Δ*recF*	0.217

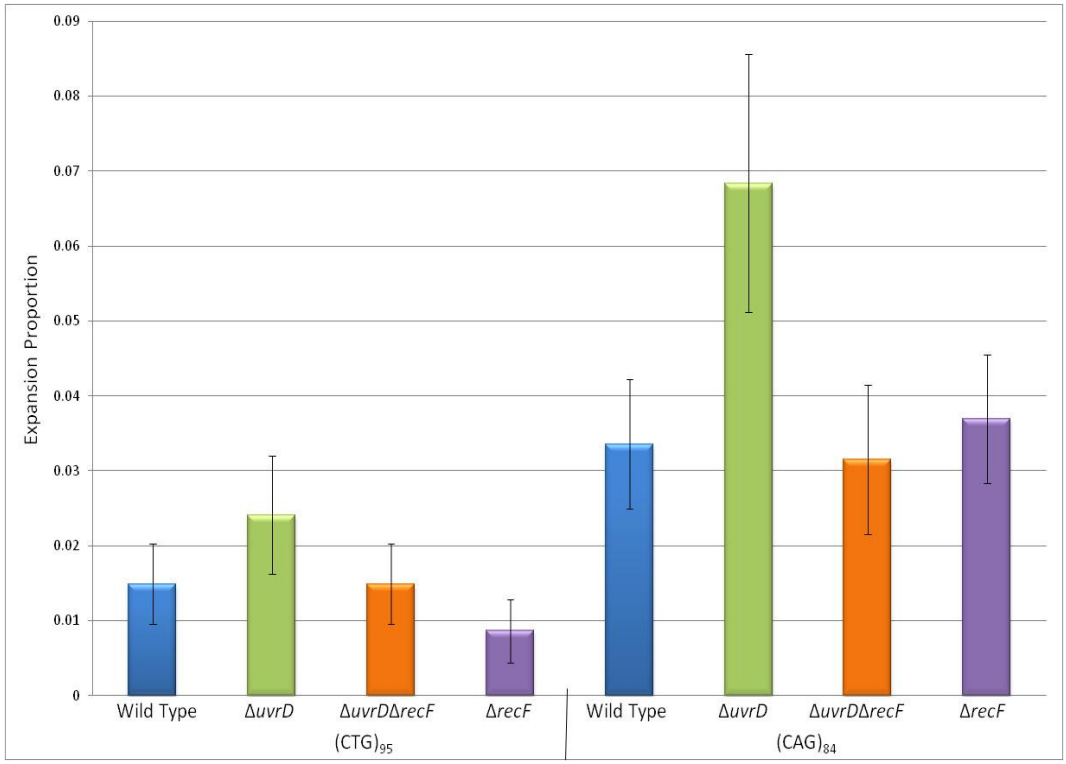

Figure 3.5: Expansion instability proportion of CTG•CAG arrays in *E. coli* cells containing a mutation in replication fork reversal genes, adjusted to include the instability occurring during growth of colonies on plate. Each bar represents the proportion of expansion events calculated from 60 independent assays including expansion events considered to have occurred during the growth of the colony on the agar plate. Error bars show the standard error of the mean calculated for each strain.

Table 3.3: Logistic regression analysis of expansion instability proportions of RFR mutant strains adjusted to include instability occurring during growth of the colony on plate. p≤ 0.05 considered statistically significant.

Orientation	Genotype	P – value
$(CTG)_{95}$	*ΔuvrD*	0.747
	ΔuvrDΔrecF	0.994
	ΔrecF	0.375
$(CAG)_{84}$	*ΔuvrD*	0.072
	ΔuvrDΔrecF	0.771
	ΔrecF	0.635

3.2.2.2 Expansion size

As the data suited a non-parametric analysis due to the low frequency of expansion lengths detected and the fact they did not form a normal distribution, the median of all the expansion lengths detected for a particular strain was calculated and indicated on a graph (Figure 3.6). While deletion lengths were found to have a bias towards large deletions, expansions have a tendency to be smaller in size (none showing a median value of over 110% of the parental TNR size).

The error bars (representing the inter-quartile range) for the $(CTG)_{95}$ wild type strain were particularly large. This result was due to the presence of one large expansion (143% parental length) and also by the very low frequency of expansion events in that particular strain, meaning that the outlying result had a strong influence on the error bars. Kruskal-Wallis analysis of the median lengths was performed (Table 3.4) to test for any significant between mutations in *uvrD* and *recF* on the size of expansion products. This analysis showed that in the strains carrying the $(CTG)_{95}$ array on the leading strand there is no significant difference between any of the median values calculated for expansion length ($p=0.319$). However, in the strain carrying $(CAG)_{84}$ on the leading strand there is a difference between expansion lengths in the mutant strains tested that is close to being statistically significant ($p=0.054$). This analysis would indicate that the Δ*uvrD*Δ*recF* strain is producing smaller length expansion products than one of the other strains, though this result does not quite achieve the 95% confidence level sought. Additionally, if it is indicative of a difference in the expansions

produced, such differences are relatively small and may not reflect a significant difference in the intermediates formed during the instability process.

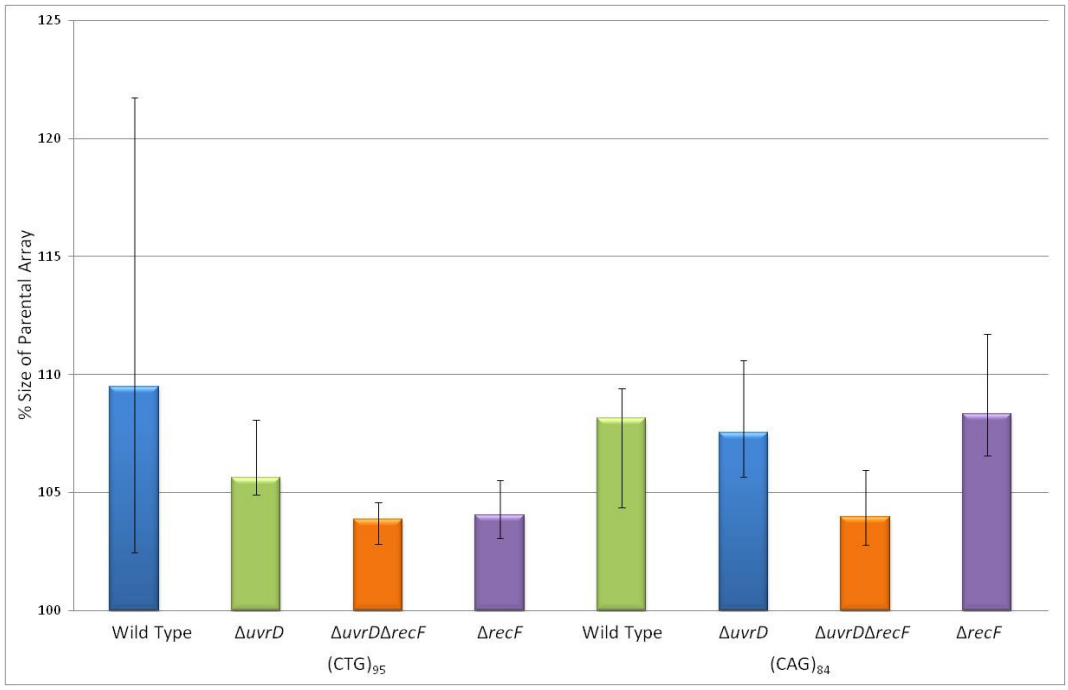

Figure 3.6: Median expansion size as a percentage of parental array length in *E. coli* strains carrying a CTG•CAG repeat array. The median expansion size was calculated for each strain from each of the expanded lengths observed. Bars represent the inter-quartile range for each median.

Table 3.4: Kruskal-Wallis analysis of median expansion lengths of RFR mutant strains. p≤ 0.05 considered statistically significant.

Orientation	Strain	Median	Rank	P – value
$(CTG)_{95}$	Wild Type	109.5	18.8	
	Δ*uvrD*	105.6	19.5	
	Δ*uvrD*Δ*recF*	103.9	11.9	
	Δ*recF*	104.0	14.0	
	Overall		17.0	0.319
$(CAG)_{84}$	Wild Type	108.1	28.2	
	Δ*uvrD*	107.5	31.1	
	Δ*uvrD*Δ*recF*	104.0	16.1	
	Δ*recF*	108.3	34.6	
	Overall		29.0	0.054

3.3 Discussion

3.3.1 Deletion instability

The study of deletion instability of CTG•CAG TNR in wild type strains confirms the previously reported observation that this type of instability is dependent on the orientation of the repeat array relative to the origin of replication (Kang *et al.*, 1995a). This is understood to be due to the differing sequence present in the transiently single stranded region of the lagging strand template during replication. The orientation in which this sequence produces a more thermodynamically stable hairpin-like structure experiences a greater proportion of deletion instability.

Deletion of the *uvrD* gene, encoding a helicase protein, has a significant effect on the deletion instability of a TNR array in both orientations of repeat sequence, resulting in an approximately 2 to 3-fold increase in instability. The UvrD protein acts to clear the RecA protein away from single stranded DNA, allowing one pathway of fork reversal to take place (Flores *et al.*, 2005). A cell deficient in protein would be accumulating RecA filaments at sites of fork arrest. Therefore a Δ*uvrD* mutant strain would be unable to accomplish replication fork reversal through this pathway. If the fork reversal pathway was involved in deletion instability of TNR, a Δ*uvrD* mutant cell would show less deletion instability than wild type cells. Furthermore, a Δ*uvrD*Δ*recF* double mutant would have a similar instability level to a wild type strain, in this mutant RecA is not loaded onto single stranded DNA so does not block fork reversal (Flores *et*

al., 2005). However, this work shows an increase in deletion instability in the Δ*uvrD* mutant and little effect of the additional Δ*recF* mutation in either orientation, a logistic regression model applied to the data to test for a difference between Δ*uvrD* and Δ*uvrD*Δ*recF* strains confirmed this (p=0.161 and p=0.261 for the $(CTG)_{95}$ and $(CAG)_{84}$ orientations respectively) . This indicated no role of RFR in deletion instability of CTG•CAG repeats.

The size of deletion instability products showed a negative skew towards larger deletions in both orientations. In the CAG leading strand orientation all strains tested displayed a similar distribution of deletion products. However, in the opposite orientation, there appeared to be a slight difference in the strains containing a *recF* mutation (Figure 3.2). Notably, this CTG leading strand orientation showed much lower frequency of deletion events compared to the CAG leading strand orientation, this apparent difference may just be an artefact of that disparity in the events measured. If the effect of *recF* mutation on the size of deletion products is real it does not produce a big difference in deletion distribution, the strains still have some negative skew. Average (median) deletion sizes varied slightly between mutants, however whether these differences are significant remains questionable. The differences in average size are small relative to the size of the array and may not reflect a difference in the intermediates produced in the instability pathways involved.

3.3.2 Expansion instability

Deletion instability is said to predominate over expansion instability in prokaryotes (Cleary *et al.*, 2002). However, Figure 3.1 and 3.4 show that for strains carrying the $(CTG)_{95}$ array on the leading strand deletion instability proportion and expansion instability proportion are almost equal. This is due to there being fewer deletion events detected in the $(CTG)_{95}$ wild type strain, as the less thermodynamically stable CAG array is present in the transiently single stranded region of the lagging strand template, making this a much more genetically stable orientation. Strains carrying TNRs in the opposite orientation do show a significant predomination towards deletion instability.

Unlike for deletion instability, the results produced in this chapter show for the first time that there is no orientation dependence on expansion instability. This result is interesting as so far all proposed models of expansion of TNR tracts are dependent on structure formation, albeit on the newly synthesized strand rather than the template strand. The fact that expansion instability seems to not be dependent on the orientation of the repeat sequence relative to the origin of replication could mean that hairpin-like structures are not forming, or that the formation of such structures is not the limiting step in the expansion instability pathway, or it may reflect an instability pathway in which expansion arises with equal chance from the either newly synthesized leading or lagging strands. An alternative explanation for the lack of orientation dependence may be that expansion instability occurs independently of replication in *E.coli*. As expansion instability requires structure formation on a newly synthesised DNA strand

DNA synthesis would still be required, but it may be that it is generated through post-replicative recombination or repair processes.

It was predicted that, if expansions were replication fork reversal dependent, strains which were unable to reverse an arrested fork would show a decrease in expansion proportion compared to a wild type strain containing a CTG•CAG TNR in the same orientation. However, expansion proportion in a Δ*uvrD* strain was not significantly different from that of the wild type strain in either orientation, nor did the double Δ*uvrD*Δ*recF* mutant alter the expansion instability proportion significantly. These results suggest that this particular pathway of replication fork reversal is not involved in the generation of expansion instability of CTG•CAG TNR in *E. coli*.

Interestingly, studies on the yeast UvrD homologue Srs2 led to different results Two separate groups showed that expansion levels increased significantly in a srs2Δ strain (Bhattacharyya and Lahue, 2004; Kerrest *et al.*, 2009). Though only one of these studies found an effect of Srs2 on deletion instability in *S. cerevisiae* (Kerrest *et al.*, 2009). However, though Srs2 and UvrD share some common features in both organisms, they may act differently in TNR instability pathways. An *in vitro* study into the ability of both proteins to unwind hairpin-like structures formed by TNR sequences showed that the Srs2 protein unwinds such a structure significantly faster than the *E.coli* UvrD protein (Bhattacharyya and Lahue, 2005). One yeast model proposed that Srs2 acts in the cell to unwind hairpins on newly synthesized DNA strands, preventing expansions

from forming (Bhattacharyya and Lahue, 2004). Other work using a longer TNR array in the yeast chromosome proposed that the Srs2 protein acts at arrested replication forks to clear Rad51, a RecA homologue, from single stranded DNA allowing fork reversal. In this model, fork reversal reduces expansion instability by preventing breakage of the DNA and expansion instability resulting form recombinational repair of this broken product (Kerrest *et al.*, 2009). In both of these models Srs2 is having a significant effect on the expansion instability of TNR arrays, which was not observed in the system used here. The second study used arrays of 70 repeat units, as opposed to arrays of 25 units or less in the first study. The repeat arrays used in the work presented here were 95 and 84 repeat units long, depending on the orientation. It may be that longer TNR arrays expand through a different mechanism to shorter arrays in yeast, one involving recombination (Kerrest *et al.*, 2009). Though no significant effect of *uvrD* deletion on TNR expansion was seen, it maybe that an effect is noticeable upon mutation of the recombination machinery (see chapter 4).

Analysis of the sizes of expansion products produced, did not detect any significant difference between any of the strains for a particular orientation. The median expansion for each strain was relatively small compared to the length of the parental repeat array. This suggests that there is no effect of replication fork reversal on the size of expansions taking place in CTG•CAG TNR arrays.

In the RFR pathway tested in this study, the presence of RecA protein on single stranded DNA inhibits the reversal mechanism (Flores *et al.*, 2005). Another RFR pathway has been identified in DnaB mutant cells, which has not

been investigated in this chapter. In this, RecA protein activates the pathway by initiating strand invasion and producing the reversed structure (Seigneur *et al.*, 2000). Therefore, *recA* mutant strain was also investigated (Chapter 4) to determine the effect of recombination proteins on expansion and deletion instability.

Mirkin and colleagues suggested that instability of a TNR array was dependent on replication fork reversal (Mirkin, 2007). However, work presented in this chapter has shown that both deletion and expansion instability of CTG•CAG TNR arrays are not dependent on the UvrD dependent pathway of replication fork reversal. While the data in this chapter ruled out this particular pathway of RFR, they also pointed to a role of the UvrD protein in the deletion instability pathways of TNRs, independent of its role in RFR. It is hypothesized that it may be the helicase function of UvrD protein which is responsible for its effects on instability, so to investigate this other helicase mutant strains were investigated (Chapter 5).

Chapter Four:

Gap Repair and Trinucleotide Repeat Instability

4.1 Introduction

This chapter describes studies performed to determine the effect of genes encoding gap repair proteins on TNR instability. Homologous recombination is the process by which both double stranded breaks and single stranded gaps are repaired in *E. coli* and other organisms (Kowalczykowski, 2000).

In human TREDs, there is little evidence of an effect of recombination on flanking markers surrounding TNR arrays (Pearson *et al.*, 2005). Studies in *E. coli* have demonstrated some effect of HR on the instability of TNR arrays *in vivo*, but the results of these studies are somewhat contradictory. Some have shown that an active HR system can lead to instability of TNR arrays (Hashem *et al.*, 2004; Jakupciak and Wells, 2000; Pluciennik *et al.*, 2002), whilst others have shown that HR works in a mechanism to stabilize TNR arrays, thus TNR instability was reduced in a HR proficient strain (Hebert *et al.*, 2004; Hebert and Wells, 2005; Sopher *et al.*, 2000). Perhaps one of the problems with these studies, leading to the discrepancy of the results, is the experimental system used. In all *E. coli* studies a plasmid based assay was utilized, which might not reflect the biology of chromosomal TNR arrays *in vivo*. In this study I have tested the effect of recombination genes that encode proteins involved in gap repair to determine what effect, if any, they have on chromosomal CTG•CAG TNR arrays.

The RecFOR complex is required for gap repair in *E. coli* (see Chapter 1.4.6). RecF is a protein capable of binding DNA, which during gap repair binds

to the ssDNA-dsDNA junction of the gap (Griffin and Kolodner, 1990; Morimatsu and Kowalczykowski, 2003). RecF protein also binds the RecR protein, to form a complex which can act to block extension of a RecA filament into regions of dsDNA (Webb *et al.*, 1997). RecO protein also forms a complex with RecR, and it is in this way RecR is believed to act as a bridge in the RecFOR complex (Morimatsu and Kowalczykowski, 2003). The function of the RecFOR complex at single stranded gaps is to displace SSB protein bound to the DNA, thus allowing RecA filament to form on the ssDNA, initiating the HR repair process (Figure 1.7) (Morimatsu and Kowalczykowski, 2003).

Here I show that mutations in genes coding for proteins involved in gap repair have no effect on the frequency of chromosomal TNR instability. Furthermore, following mutation in any of the genes tested, no effect was seen on the size of expansion products produced in either orientation of the TNR array. The size of deletion products in strains carrying TNR in the CAG orientation was also unaffected by a mutation in any of the gap repair genes tested. However, an effect was detected on the size of deletion product in strains carrying a CTG TNR array on the leading strand.

4.2 Results

4.2.1 Deletion instability in gap repair mutants

4.2.1.1 Deletion instability proportion

The mean deletion proportion was calculated from 60 independent overnight cultures for each strain investigated, and plotted together in Figure 4.1. A logistic regression model was applied to the data, to determine if any of the mutant strains investigated deviated significantly from the wild type control for that orientation of the TNR array (Table 4.1).

As seen in the previous chapter, deletion proportion in the wild type and all mutant strains tested was dependent on the orientation of the TNR array relative to the origin of replication of the chromosome. The strains investigated had a deletion proportion 5 to 6-fold higher in cells containing the TNR in the CAG orientation, in which the thermodynamically stable CTG sequence was present on the lagging strand template.

For each orientation, comparison of mutant strains to the wild type showed that there was no significant difference in deletion proportion between any of the gap repair mutants studied and the control strains (Figure 4.1). This negative result was confirmed statistically by logistic regression analysis of the data (Table 4.1). None of the mutant strains in either orientation showed a statistically significant difference from the control strain.

Table 4.1 Logistic regression analysis of deletion instability proportions of gap repair mutant strains. p≤ 0.05 considered statistically significant.

Orientation	Strain	P-Value
$(CTG)_{95}$	Δ*recF*	0.659
	Δ*recO*	0.667
	recA$^-$	0.235
$(CAG)_{84}$	Δ*recF*	0.183
	Δ*recO*	0.893
	recA$^-$	0.709

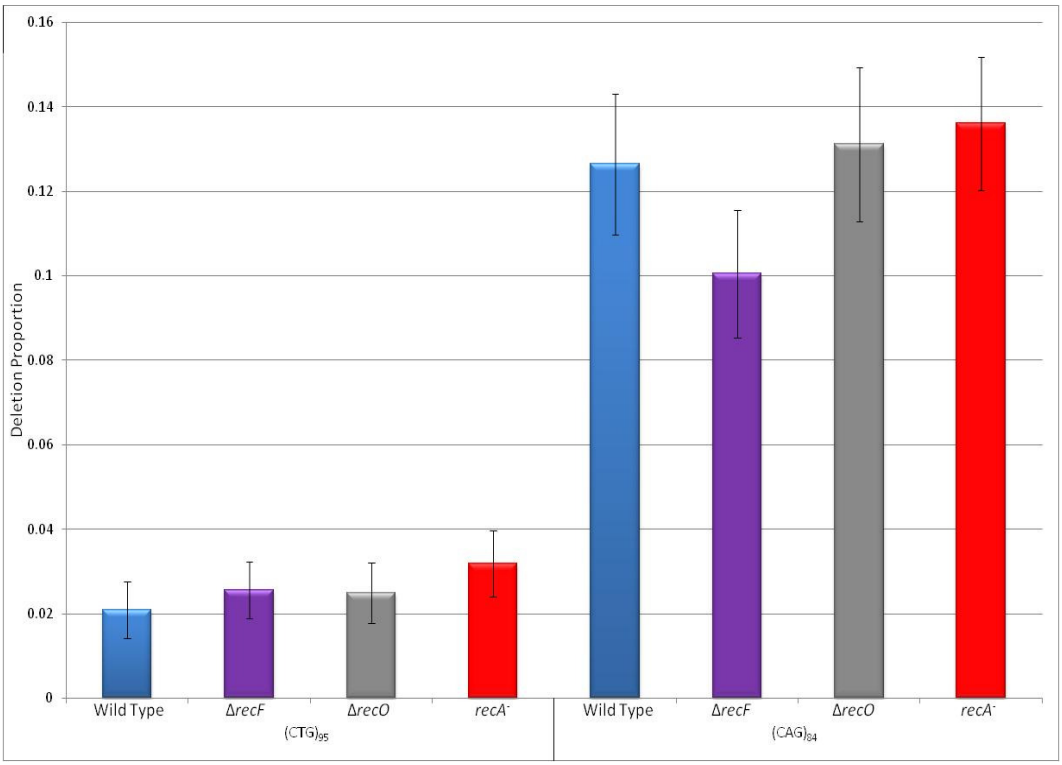

Figure 4.1: Deletion instability proportion of CTG•CAG arrays in *E. coli* strains containing a mutation in genes involved in the gap repair pathway. Each bar represents the proportion of deletion instability events calculated from 60 independent assays. Error bars show the standard error of the mean calculated for each strain.

4.2.1.2 Deletion size

Detected deletion lengths, both those included in the instability assay and those considered to have been produced during growth of the colony on the agar plate, were converted to a percentage size of the parental TNR array length, and the frequency of each counted. These data were used to produce a figure showing the deletion distribution for each strain tested (Figure 4.2 and Figure 4.3).

Figure 4.2 showed that strains carrying $(CAG)_{84}$ TNR on the leading strand all displayed a very similar deletion length distribution. The three mutant strains produced a similar negative skew to that seen in the wild type strain for that orientation, suggesting the gap repair genes are not involved in the deletion pathway of TNR arrays in this orientation. This was confirmed by calculating the average (median) size of deletion product. Mutant strains in this orientation all displayed a median very similar to the wild type control. Though similar distribution patterns were seen, the area underneath the trend line was lower for the $(CAG)_{84}$ *recA*$^-$ strain. Figure 4.1 shows that instability in the growth of the overnight culture was the same for all four $(CAG)_{84}$ strains tested. The area underneath the trend line should be proportional to the frequency of deletion events in that strain, suggesting that the difference between the four strains is due to deletion events occurring during the growth of the colony on a plate.

Figure 4.3 however, showed that the mutations in a strain carrying TNR arrays in the CTG orientation did have an effect on the size of deletion product produced. It was noted in the previous chapter that a Δ*recF* mutant in this orientation showed a slightly less negative skew to the distribution of

deletion products. The effect of this mutation was only slight so whether it reflected a real difference in the deletion pathways involved in this mutant was questioned. The results presented in this chapter, of strains carrying TNR arrays in the same orientation with mutation in the *recO* and *recA* genes, supported this initial observation in the *recF* mutant. Both *recO* and *recA* mutant strains showed a similar deletion distribution pattern to the *recF* mutant, different to the wild type control for this particular orientation of TNR array. While the effect was relatively small, it was certainly repeatable within this group of mutants, all involved in the gap repair pathway. Again this was also evident in the average (median) size of deletion product, all the mutant strains tested produced smaller deletion products than the wild type control.

The deletion distribution graphs of strains in the CTG orientation indicated that there was a slightly different distribution of deletion products in the gap repair mutants then the wild type strain. The mutant strains tested still produced a negatively skewed distribution of deletion products, although less negatively skewed than the wild type.

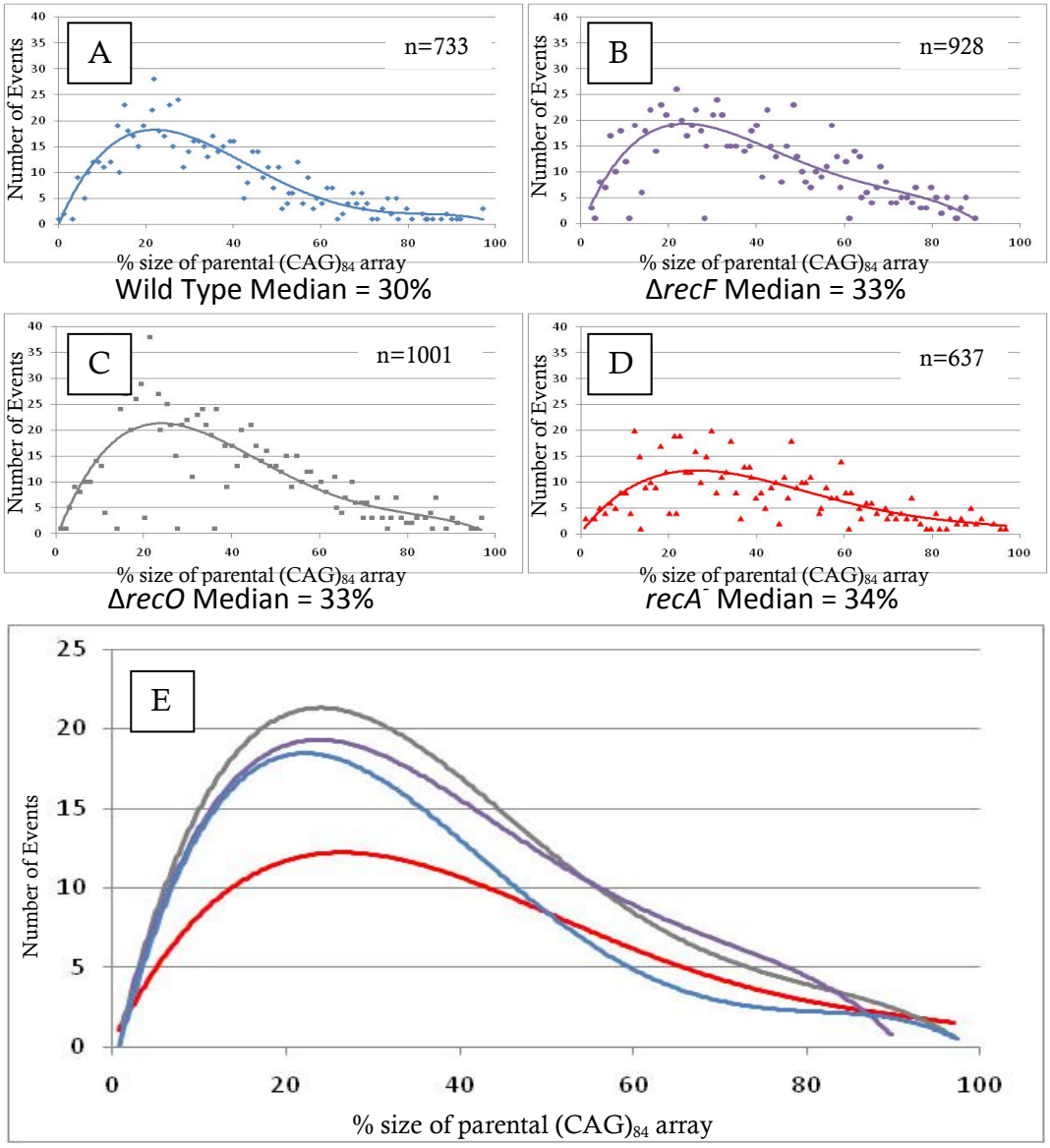

Figure 4.2: Distribution of deletion sizes of *E. coli* strains carrying the $(CAG)_{84}$ repeat array on the leading strand. Deletions are presented as a percentage of the parental 84 repeat array plotted against the frequency of the observed deletion size for strains: Wild Type (**A**), Δ*recF* (**B**), Δ*recO* (**C**) and *recA*$^-$ (**D**). Trend lines for each distribution were plotted together (**E**).

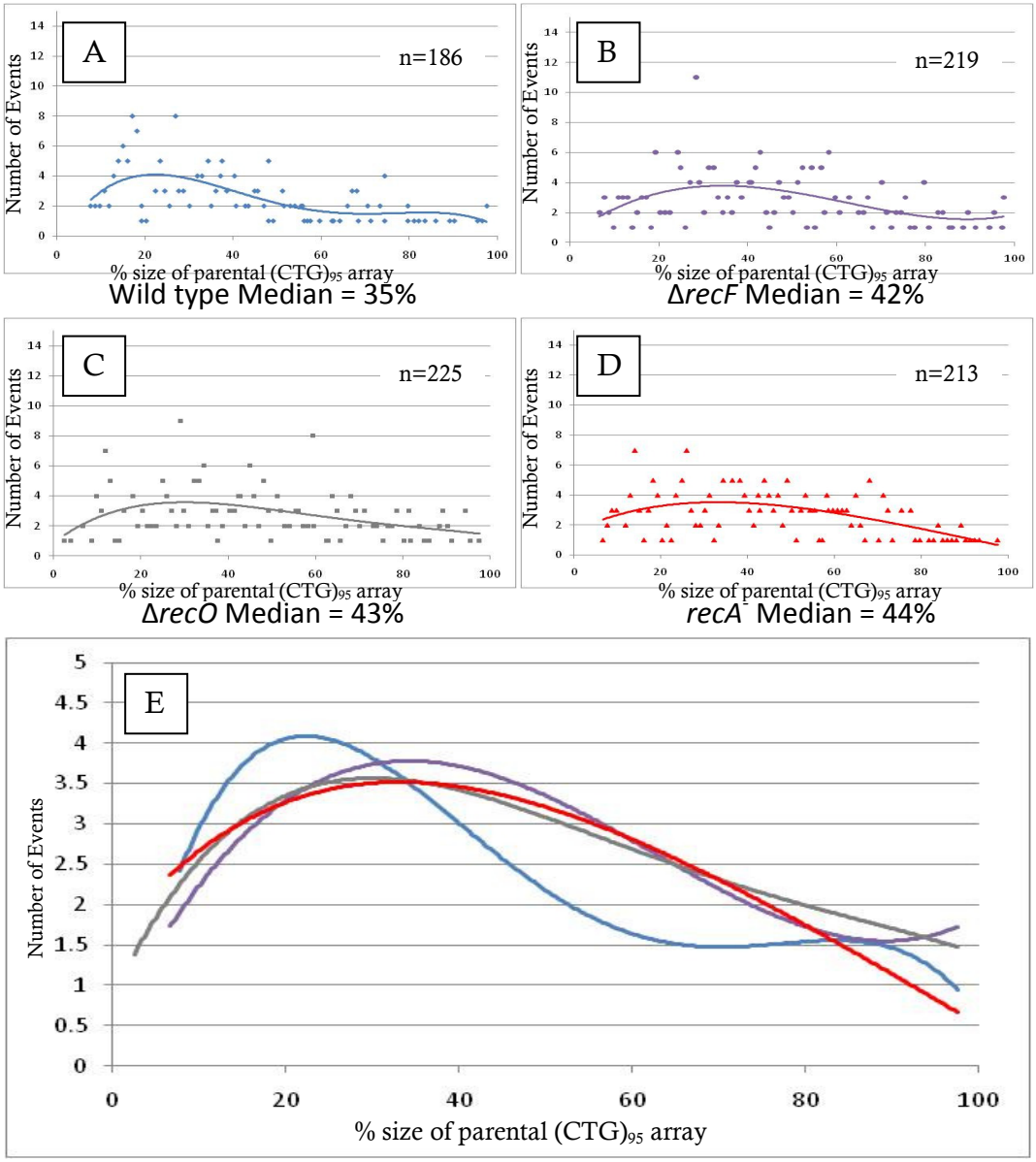

Figure 4.3: Distribution of deletion sizes of *E. coli* strains carrying the $(CTG)_{95}$ repeat array on the leading strand. Deletions are presented as a percentage of the parental 95 repeat array plotted against the frequency of the observed deletion size for strains: Wild Type (**A**), Δ*recF* (**B**), Δ*recO* (**C**) and *recA*$^-$ (**D**). Trend lines for each distribution were plotted together (**E**).

4.2.2 Expansion instability in gap repair mutants

4.2.2.1 Expansion instability proportion

Expansion proportion was calculated from 60 independent overnight cultures and an average for each strain was determined and plotted in Figure 4.4.

The results for presented in Figure 4.4 showed that expansion instability was orientation independent in the *recF, recO* and wild type strains but less in the *recA*$^-$ mutant strain. Though there appeared to be an orientation effect on expansion proportion in this *recA*$^-$ mutant, the size of the error bars and the relatively small difference between orientations cast doubt on whether this reflected a true difference in the mechanisms involved.

Logistic regression analysis of the data (Table 4.2) confirmed that the mutant strains tested within a particular orientation did not significantly differ from the wild type strains carrying TNR arrays in the same orientation.

As the results presented in Figure 4.4 are less clear than those of the previous chapter the alternative analysis method, including those expansion events considered to have occurred during growth of the colony on agar plate was also conducted (Figure 4.5). In the previous chapter this analysis was complicated by the fact that deletion instability varied between strains carrying a TNR in a particular orientation. As this alternative analysis precluded those colonies in which a deletion event had occurred during the overnight culture growth, these differences were reflected in the analysis. However, as Figure 4.1 showed, deletion instability is not significantly affected by a mutation in one of

the gap repair genes between strains carrying a TNR array in the same orientation, meaning this would be less of an issue in these strains. Logistic regression analysis was also conducted on these data to determine if any of the strain differed from the wild type control for that particular orientation of TNR array (Table 4.3).

The results of the alternative analysis confirmed that the gap repair genes have no effect on expansion instability in strains carrying TNR arrays in either orientation. The differences seen in the original analysis were less obvious in the alternative analysis and the variation within each strain was reduced, as the error bars were smaller in the second analysis. The results presented in Figure 4.5 appear to show an orientation dependence, however this is an artefact of the analysis not including colonies in which a deletion event had taken place in the growth of the overnight culture. As deletions show an orientation dependence, then this will be reflected in the alternative analysis. Logistic regression analysis confirmed this result, that none of the strains tested had a statistically significant effect on the expansion proportion seen within that strain.

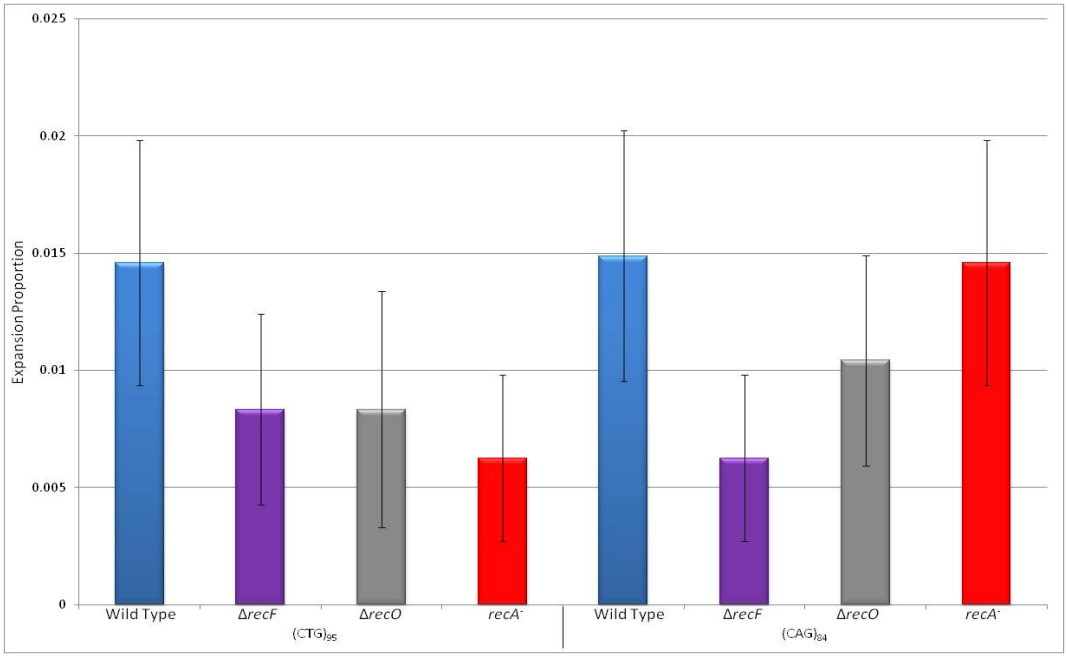

Figure 4.4: Expansion instability proportion of CTG•CAG arrays in *E. coli* strains containing a mutation in genes involved in gap repair pathways. Each bar represents the proportion of expansion instability events calculated from 60 independent assays. Error bars show the standard error of the mean calculated for each strain.

Table 4.2: Logistic regression analysis of expansion instability proportions of strains mutated in one of the gap repair genes. p≤ 0.05 considered statistically significant.

Orientation	Strain	P-Value
$(CTG)_{95}$	Δ*recF*	0.373
	Δ*recO*	0.116
	recA$^-$	0.218
$(CAG)_{84}$	Δ*recF*	0.217
	Δ*recO*	0.570
	recA$^-$	0.777

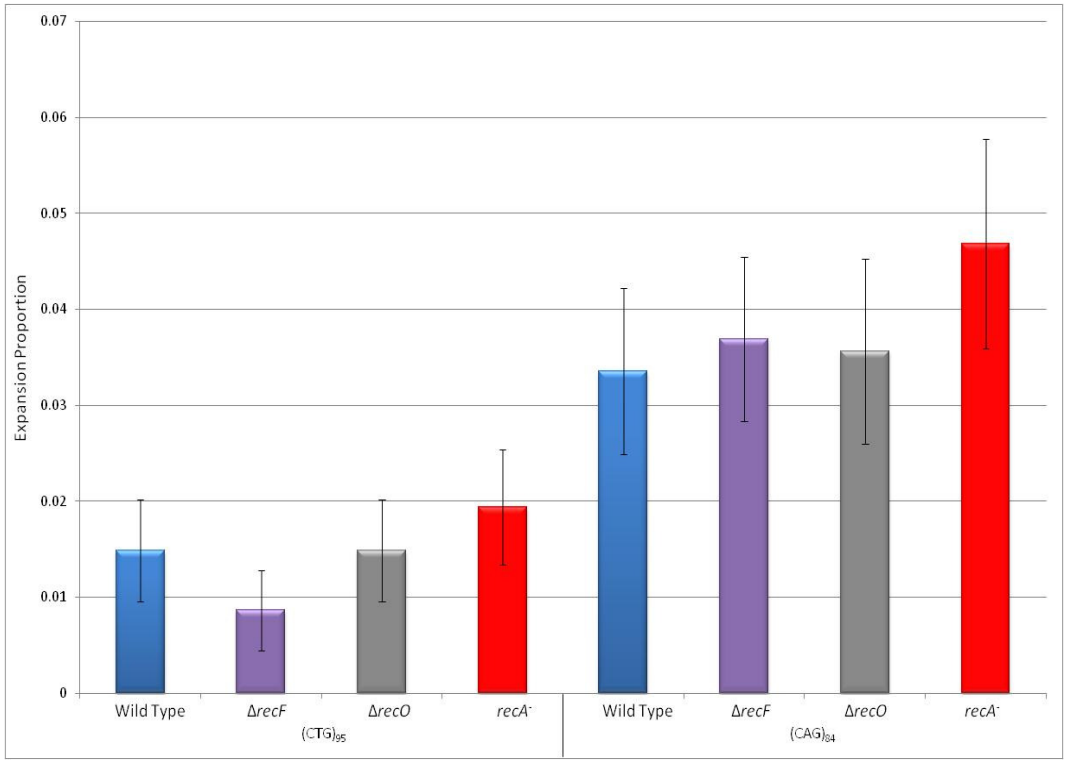

Figure 4.5: Expansion instability proportion of CTG•CAG arrays in *E. coli* cells containing a mutation in gap repair genes, adjusted to include the instability occurring during growth of colonies on plate. Each bar represents the proportion of expansion events calculated from 60 independent assays including expansion events considered to have occurred during the growth of the colony on the agar plate. Error bars show the standard error of the mean calculated for each strain.

Table 4.3: Logistic regression analysis of expansion instability proportions of gap repair mutant strains, adjusted to include instability occurring during growth of the colony on plate. p≤ 0.05 considered statistically significant.

Orientation	Genotype	P – value
$(CTG)_{95}$	Δ*recF*	0.375
	Δ*recO*	0.990
	recA$^-$	0.594
$(CAG)_{84}$	Δ*recF*	0.635
	Δ*recO*	0.549
	recA$^-$	0.438

4.2.2.2 Expansion size

Average expansion size was studied using a non-parametric analysis method again, as there was a low frequency of expansion events detected. Median expansion size was calculated for each strain and plotted (Figure 4.6). Kruskal-Wallis analysis was carried out to test if there was any statistically significant difference between any of the strains containing a particular orientation of the TNR array (Table 4.4).

The variation within some of the strains was quite large, particularly in the $(CTG)_{95}$ wild type strain and *recA*$^-$ strains in both orientations. However, the median data showed that none of the mutant strains had an effect on the expansion size detected. All strains showed relatively small median expansion sizes – less than 110% of the parental array size, in contrast to the deletion distributions which, even in the gap repair mutants, show a bias towards big deletion events. This was confirmed by the Kruskal-Wallis analysis which found no statistically significant difference in either orientation of the TNR array ($p= 0.264$ and $p= 0.181$ for $(CTG)_{95}$ and $(CAG)_{84}$ respectively).

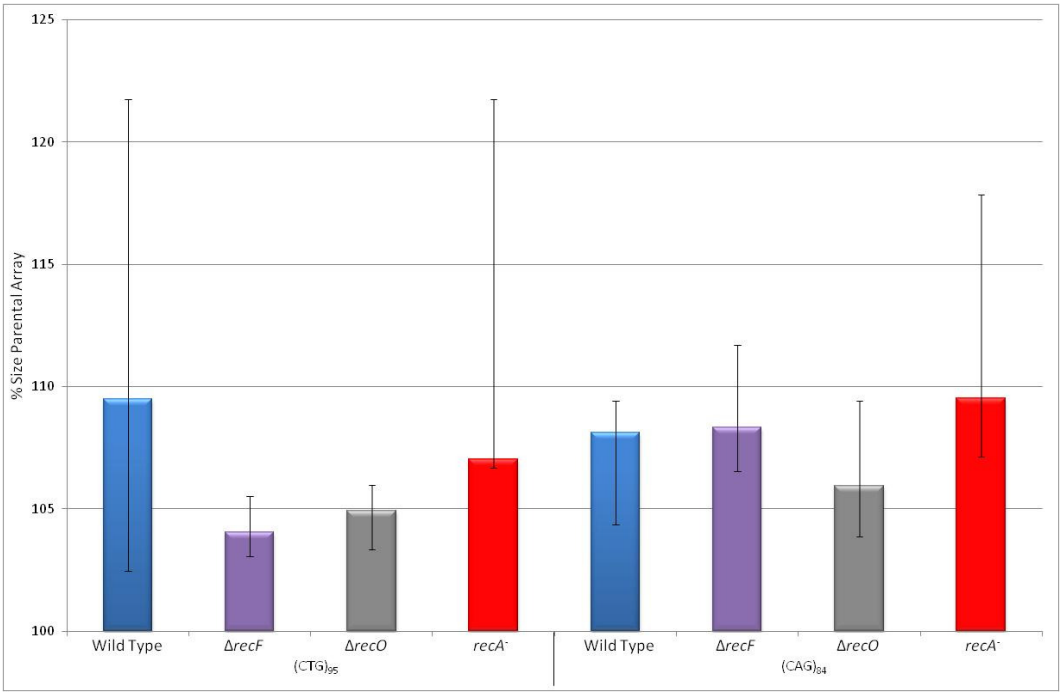

Figure 4.6: Median expansion size as a percentage of parental array length in *E. coli* strains carrying a CTG•CAG repeat array. The median expansion size was calculated for each strain from each of the expanded lengths observed. Error bars represent the inter-quartile range for each median.

Table 4.4: Kruskal-Wallis analysis of median expansion lengths of strains mutated in one of the gap repair genes. p≤ 0.05 considered statistically significant.

Orientation	Strain	Median	Rank	P – value
$(CTG)_{95}$	Wild Type	109.5	16.0	
	Δ*recF*	104.0	12.0	
	Δ*recO*	104.9	10.8	
	recA$^-$	107.0	18.6	
	Overall		15.0	0.264
$(CAG)_{84}$	Wild Type	108.1	25.3	
	Δ*recF*	108.3	31.3	
	Δ*recO*	106.0	21.4	
	recA$^-$	109.5	34.4	
	Overall		29.0	0.181

4.3 Discussion

Studies into the effect of HR on TNR instability have provided mixed, often conflicting, results in the past (Hashem *et al.*, 2004; Hebert *et al.*, 2004; Hebert and Wells, 2005; Jakupciak and Wells, 2000; Pearson *et al.*, 2005; Pluciennik *et al.*, 2002). Those studies conducted in *E. coli* have always utilized plasmid based systems, so the question remained as to the effect of HR on chromosomally based TNR arrays.

4.3.1 Deletion instability

Mutation of genes encoding gap repair proteins was found to have no effect on CTG•CAG TNR deletion instability proportion for strains carrying the TNR array in either orientation relative to the origin of replication. This result suggests that gap repair is not involved in the deletion instability pathway. The fact that a *recA* mutation did not affect deletion proportion, compared to a wild type strain, also suggests that recombination in general does not play a role on the deletion instability in the strains studied.

Analysis of the deletion sizes produced showed that in strains carrying CAG TNR arrays on the leading strand instability was not affected by mutation of the gap repair genes. Strains in which the CTG array was present on the leading strand, however, were affected by the mutation of the gap repair genes. In all the mutants tested the distribution of deletion products were less biased towards big deletions than in the wild type strain carrying the TNR in the same

orientation (Figure 4.2), confirming the effect of a *recF* mutation seen in the previous chapter. This suggests that the gap repair genes, whilst not affecting the number of deletion events occurring, do affect the size of the deletion product produced, suggesting they may play a role in the processing of the intermediates formed in the deletion pathway of strains carrying a TNR in this orientation (for further discussion see Chapter 5).

4.3.2 Expansion instability

This study showed that, like deletion instability, there was no significant effect of the gap repair genes on expansion proportion of TNR arrays, irrespective of the orientation of the TNR array relative to the origin of replication. The result, particularly that of the *recA*$^-$, suggests no role for recombination pathways on the expansion of TNR arrays in the *E.coli* chromosome.

Analysis of the size of expansion products showed no significant effect on expansion sizes detected in any of the gap repair mutant strains. While the mutant strains had some effect on deletion size of instability event occurring in strains carrying CTG TNR array on the leading strand, the same was not true for expansion sizes. This suggests that there are different intermediates and pathways of instability involved in deletion and expansion instability.

Previous studies into the effect of HR on the instability of TNR arrays have produced conflicting results, some suggesting a role for HR in maintaining stability of a TNR array, while others propose HR can lead to instability. The results in this chapter suggest that, for a TNR array in the *E. coli* chromosome HR has no effect on instability. If HR were the cause of TNR instability we would expect less instability in a HR deficient mutant strain, if it were preventing instability we would expect more in a HR deficient mutant. No effect on the proportion of deletion or expansion instability events was seen in HR deficient mutant strains, suggesting no role for the HR pathway in TNR instability.

Chapter Five:

Helicases and Trinucleotide

Repeat Instability

5.1 Introduction

Results presented in Chapter 3 showed that strains in which the *uvrD* gene had been deleted displayed a higher level of deletion instability. UvrD protein was shown to act by preventing instability events *in vivo*, and it was suggested that it may be the helicase activity of the protein responsible for this. This confirmed results seen in the *Saccharomyces cerevisiae* homologous protein Srs2 (Kerrest *et al.*, 2009). To investigate this further a number of strains mutated in genes encoding helicase proteins were assayed for their instability proportions.

Duplex DNA stores information held within a cell. This duplex must be unwound to release the information allowing important cellular processes to occur. For this purpose a class of proteins called helicases transiently unwind the DNA, permitting access to the information stored within (Tuteja and Tuteja, 2004). Helicases are ubiquitous proteins, which actively separate hydrogen bonds between DNA duplexes or DNA-RNA hybrids by utilising energy from hydrolysis of NTP. They are important therefore, in many pathways within the cell including replication, recombination, repair, transcription and translation (Singleton *et al.*, 2007).

Helicase proteins bind to DNA, many by recognising particular structures adopted by the nucleic acid. *E. coli* helicases have been identified binding to a variety of DNA structures including single stranded DNA, double stranded DNA, Holliday junctions and blunt-ended DNA (Gupta *et al.*, 2010; Taylor and Smith, 1985; Whitby *et al.*, 1994). Once bound to DNA, helicases have a specific

polarity of translocation along the strand, either 5'→3 or 3'→5' (Singleton *et al.*, 2007).

Organisms encode numerous different helicases present in their cells to deal with the variety of different pathways that require helicase action. *E. coli* have at least 14 different helicases including DnaB, UvrB, PriA, the RecBCD complex, helicase I, helicase III, helicase IV, Rho, RuvAB, UvrD, RecQ, RecG, DinG and Rep (Tuteja and Tuteja, 2004). Helicase proteins have been subdivided into six so-called superfamilies dependent on their primary structure (Singleton *et al.*, 2007).

In this chapter the genetic assay described previously was used to determine the effect of particular *E. coli* helicase proteins on the instability of chromosomal CTG•CAG TNR arrays *in vivo* by constructing strains carrying mutations in particular helicase genes. The role of the proteins encoded by each of these genes is described below. The role of the helicase UvrD was discussed previously in Chapter 3.

5.1.1 Rep Helicase

Initial experiments on the Rep protein revealed that chromosome replication in strains carrying a mutation in the *rep* gene was two times slower than in control wild type strains (Lane and Denhardt, 1975). This suggested a role for the protein in the replication pathway, which was supported by the observation that *rep* mutant strains undergo RFR (Seigneur *et al.*, 1998). The Rep helicase protein

also has a role in the cell in the PriC-Rep pathway of replication restart (Chapter 1.4.4) (Heller and Marians, 2005b).

The Rep protein is a member of the superfamily 1 helicases, and shares around 40% identity with the UvrD protein (Boubakri *et al.*, 2010). Like the UvrD protein it translocates along DNA in a 3'→5' direction. A *rep uvrD* double mutant strain is not viable when grown in rich media conditions, though it survives in minimal media (Guy *et al.*, 2009). This lethality is suppressible by the introduction of a third mutation in one of the genes encoding the RecFORQJ proteins (Lestini and Michel, 2008; Petit and Ehrlich, 2002). This may suggest that the lethality of the double mutant is due a replication fork stall in a *rep* mutant allowing formation of a toxic RecA nucleoprotein filament loaded onto ssDNA by the RecFOR proteins. UvrD would be required to clear such a structure. However, recent studies by McGlynn and colleagues found mutation of *recFORQJ* genes only provided a limited suppression of lethality in the *rep uvrD* mutant, suggesting that the presence of a nucleoprotein filament is not the only cause of death in this strain (Guy *et al.*, 2009).

A further activity of the Rep protein *in vitro* is its ability to remove proteins bound to DNA, which could possibly provide a block to replication (Guy *et al.*, 2009; Yancey-Wrona and Matson, 1992). This activity has also been demonstrated *in vivo,* as Rep protein is capable of removing RNA polymerase protein when encountered in a head-on collision between replication and transcription (Boubakri *et al.*, 2010). This finding is also supported by the

observation that Rep protein interacts physically and functionally with the replicative helicase DnaB, after it has been loaded onto the chromosome by the DnaC protein (Guy *et al.*, 2009) suggesting it moves with the replication fork and its inactivation could present a challenge to the unwinding activity of DnaB.

5.1.2 RecG

The protein encoded by the *recG* gene is another 3'→5' helicase. It belongs to the superfamily 2 helicases (Briggs *et al.*, 2004). The RecG helicase was first identified in a screen to find mutations producing deficiencies in recombination in *E. coli*, and the protein therefore plays a role in recombination within the cell, along with being involved in other pathways including replication and RFR (Gregg *et al.*, 2002; McGlynn and Lloyd, 2001).

In vivo experiments on *recG* mutants showed that strains carrying a second mutation in the gene encoding the replication restart protein PriA grew poorly, more than strains carrying a single mutation in either gene (Gregg *et al.*, 2002). This potential role with PriA indicates that RecG protein might be involved in a replication restart pathway.

RecG protein also has a role in processing of joint DNA molecules, *in vitro* it is able to catalyse the migration of Holliday junctions (Lloyd and Sharples, 1993), and it can act at arrested replication forks, translocating along both template strands, pulling the DNA back together and stripping off the

nascent DNA which can then anneal together forming a Holliday junction (McGlynn and Lloyd, 2001). The RuvABC proteins are also involved in processing Holliday junctions. A strain mutated in both *ruvABC* and *recG* showed an increased sensitivity to damage caused by UV light, suggesting that both work in overlapping pathways (Lloyd, 1991). The *in vitro* role of RecG helicase in processing of stalled replication forks, and the preferential binding of the protein to forks with no nascent leading strand has led to a proposition of RecG being involved in replication fork reversal *in vivo* (McGlynn and Lloyd, 2001). This proposed role is supported by indirect evidence from the organism *Bacillus subtilis,* in which *in vivo* experiments show that RecG protein co-localizes with RecQ and PriA at SSB (single strand binding) protein (Lecointe *et al.*, 2007).

5.1.3 RecQ

RecQ is also a 3'→5' helicase and a member of the superfamily 2 group of helicases. RecQ-like proteins have been found in all branches of life, including Sgs1 in *Saccharomyces cerevisiae*, and five exist in humans including BLM, WRN and RecQ4. Mutation of genes encoding the human proteins has been linked to several diseases resulting from genome instability (Zhang *et al.*, 2006). Most of the RecQ family of helicase proteins share a number of conserved regions including a helicase domain, a RecQ carboxyl terminal (RQC) domain, and some members also contain a helicase and RNaseD C-terminal (HRDC) domain

(Chu and Hickson, 2009). However, outside the conserved regions, the family members have very little homology (Bachrati and Hickson, 2003).

RecQ protein acts as a monomer capable of binding to a variety of different DNA structures, preferentially binding single stranded DNA (Zhang *et al.*, 2006). It is able to bind forked structures, such as arrested replication forks, displaying a preference for forks presenting a gap in the leading strand, as it binds to 3'-overhanging DNA better than to a 5'-overhanging DNA (Hishida *et al.*, 2004).

RecQ helicases have been shown to be involved in a number of important cellular pathways, and have been called 'genomic caretakers' as they maintain genomic stability (Chu and Hickson, 2009). The bacterial RecQ protein was first discovered by Nakayama and colleagues as a member of the RecF-pathway of double strand break repair (Nakayama *et al.*, 1985). The protein also has a role in the repair of double strand breaks by homologous recombination (HR). *E. coli in vitro* experiments have shown that RecQ can have both pro- and anti-recombinogenic effects. The helicase can facilitate HR by promoting strand invasion of a RecA filament forming a D-loop (Harmon and Kowalczykowski, 1998). However the same study also showed that the RecQ helicase, *in vitro*, is capable of disrupting this strand invasion structure (Harmon and Kowalczykowski, 1998). A similar result was obtained when illegitimate recombination between a phage particle and the bacterial genome was studied (Hanada *et al.*, 1997). Illegitimate recombination between the phage and the

bacterial chromosome can occur in short regions of homology. Strains carrying a mutation in the *recQ* gene displayed increased levels of such recombination events, indicating that the protein in wild type cells acts to prevent them. The RecQ protein may act by separating the joint molecule created between the phage and the bacterial chromosomal DNA. The dual role of the RecQ protein to both promote and restrict recombination has also been shown in the human BLM protein (Bugreev *et al.*, 2007).

Experimental work into the effect of the *Saccharomyces cerevisiae* RecQ homologue protein Sgs1 on TNR instability found that mutation of the gene encoding the yeast protein had a significant effect on deletion instability (Kerrest *et al.*, 2009). However, both this study and another conducted by a separate group (Bhattacharyya and Lahue, 2004) found that mutation of the *sgs1* gene had no effect on the expansion instability of TNR.

5.1.4 DinG

The gene encoding the DinG protein was first identified in a screen to find promoters regulated by SOS induction in the cell, and as such the expression of this gene is damage-inducible (Lewis *et al.*, 1992). DinG helicase is a member of the superfamily 2 helicases and, unlike the other helicase proteins studied in this chapter, translocates along DNA in a 5'→3' direction (Voloshin *et al.*, 2003). The DinG protein is related to helicases found in other organisms including

XPD and BACH1/FANCJ in humans, Rad15 in *Schizosaccharomyces pombe* and Chl1 and Rad3 in *Saccharomyces cerevisiae* (Koonin, 1993).

In vitro analysis of the DinG protein discovered that while it is able to unwind both DNA duplexes and DNA/RNA hybrids, it is a structurally specific helicase, with preferences for branched structures (Voloshin and Camerini-Otero, 2007). *In vivo* experimentation showed that mutation of the *dinG* gene did not affect the fork reversal pathway in *dnaE*ts or *dnaN*ts mutants (Flores *et al.*, 2004). However DinG was shown to function *in vivo* in the removal of R-loops, formed during transcription, encountered by a replication fork (Boubakri *et al.*, 2010) confirming an observation made previously *in vitro* (Voloshin and Camerini-Otero, 2007).

Strains carrying mutations in the 3'→5' helicase genes *rep*, *recG*, *recQ* and *uvrD* as well as the 5'→3' helicase gene *dinG*, were constructed to test their effect on both deletion and expansion instability. Mutation of the *rep* gene was found to have an effect on deletion instability in one orientation of the TNR array relative to the origin of replication. However, expansion instability was found to be affected by mutation in this gene in both orientations. The RecQ protein was also identified as playing an important role in expansion instability in the orientation in which CAG repeats were present on the leading strand template.

These results were further dissected to try to identify the pathways involved by studying a *priC* mutant strain, to test the *rep*/*priC* restart pathway; and by construction of double mutants. Though no role for the restart pathway

was discovered, analysis of double mutants led to the distinction of two pathways of TNR expansion in *E.coli*, one dependent on RecQ protein and one independent of this helicase.

5.2 Results

5.2.1 Deletion instability in helicase mutants

5.2.1.1 Deletion instability proportion

Deletion instability proportion was calculated from 60 independent overnight cultures of each of the helicase mutant strains studied, and the results were plotted in Figure 5.1. Logistic regression analysis was carried out on the data to determine if the instability in the mutant strains tested differed significantly from the wild type strain for that particular orientation of the repeat array (Table 5.1).

As shown previously the results in Figure 5.1 indicated that for most of the strains studied deletion instability was dependent on the orientation of the repeat array relative to the origin of replication. However, one mutant strain tested was not orientation dependent. In the $(CTG)_{95}$ leading strand orientation the Δ*rep* strain showed an increase in deletion proportion from wild type ($p<0.001$). However, in the opposite orientation of repeat sequence, $(CAG)_{84,}$ it did not ($p = 0.289$). Interestingly, deletion proportion of *rep* mutant cell in the

CTG orientation was very similar to the level of instability in the CAG orientation, removing any orientation dependence seen in the wild type strains.

As discussed in Chapter 3, the Δ*uvrD* strains showed an approximate 2-fold increase in deletion instability compared to the wild type strains, in both orientations of the TNR array. As this change in instability is not due to the role of the UvrD protein in replication fork reversal, it may instead be a consequence of its role as a helicase in the cell.

Logistic regression analysis demonstrated that none of the other mutant strains tested had a statistically significant effect on the deletion instability detected in either orientation of the repeat array (Table 5.1).

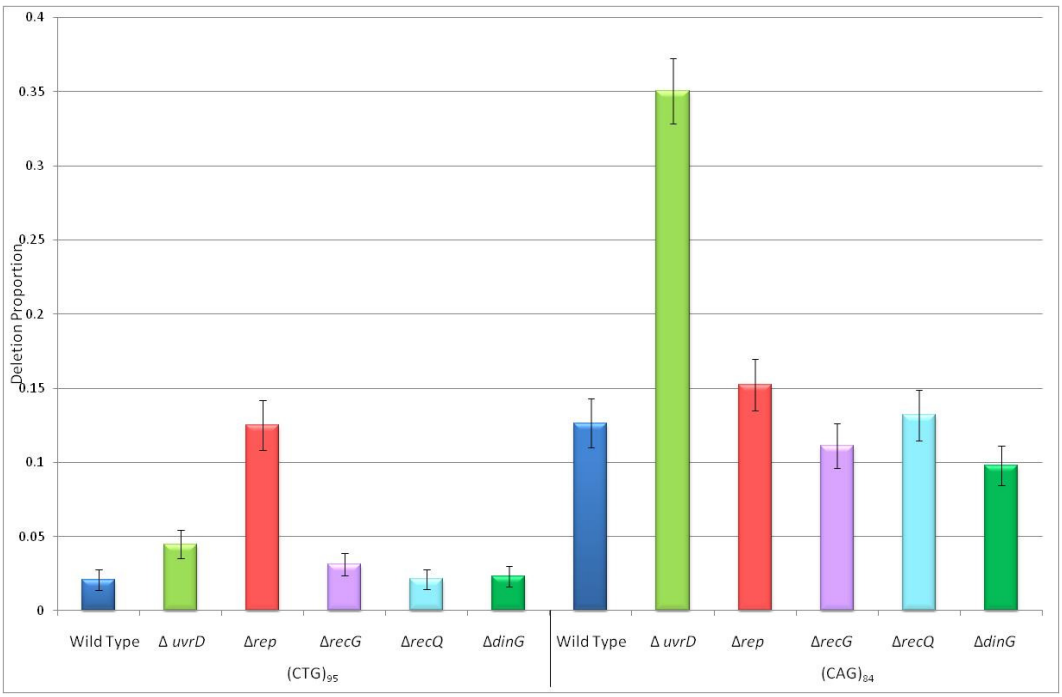

Figure 5.1: Deletion instability proportion of CTG•CAG arrays in *E. coli* strains containing a mutation in genes encoding helicase proteins. Each bar represents the proportion of deletion instability events calculated form 60 independent assays. Error bars show the standard error of the mean calculated for each strain.

Table 5.1: Logistic regression analysis of deletion instability proportions of helicase mutant strains. p≤ 0.05 considered statistically significant.

Orientation	Strain	P-Value
$(CTG)_{95}$	Δ*uvrD*	0.048
	Δ*rep*	<0.001
	Δ*recG*	0.314
	Δ*recQ*	0.996
	Δ*dinG*	0.825
$(CAG)_{84}$	Δ*uvrD*	<0.001
	Δ*rep*	0.289
	Δ*recG*	0.401
	Δ*recQ*	0.913
	Δ*dinG*	0.109

5.2.1.2 Deletion size

Deletion distribution graphs were plotted by converting all deletion lengths observed, both those thought to have been generated during growth in the overnight culture and those from growth of the colony on the agar plate, to a percentage size of the parental repeat array. Average (median) deletion sizes were calculated, and these data were displayed in Figures 5.2 and 5.3.

Strains carrying TNR arrays in both the $(CTG)_{95}$ and $(CAG)_{84}$ orientations, mostly showed the characteristic negative skew previously observed. In these strains there is a predominance of large deletions.

In both orientations of repeat array the Δ*rep* strain had a much higher frequency of instability events than any of the other strains tested. In the $(CTG)_{95}$ orientation this is consistent with the Δ*rep* strain having the highest level of deletion instability. However in the $(CAG)_{84}$ orientation the Δ*uvrD* strain had a higher deletion proportion, but a higher frequency of deletion products was

detected in the Δ*rep* mutant, which meant there must have been more instability events occurring during growth of the colony on the agar plate in this mutant.

In the $(CTG)_{95}$ orientation the trend line fitted to the Δ*recG* strain is much flatter than those observed for any of the other strains, is does not appear to display the negative skew the others do. This maybe an indication of differing intermediates produced during the instability pathway in this mutant. This distribution of deletion products is not seen in the equivalent $(CAG)_{84}$ strain to such an obvious extent.

As median values could not be compared statistically, it is difficult to draw conclusions from the slight differences in median values in some of the strain. However, strains carrying CTG arrays on the leading strand template did show a big difference in the average size of deletion product when the *recG* gene was mutated (59% compared to 35%). This indicated that smaller deletion products were produced in the strain mutated in the *recG* gene, suggesting that RecG protein affects the size of deletion events occurring but not the frequency of such events (Figure 5.1).

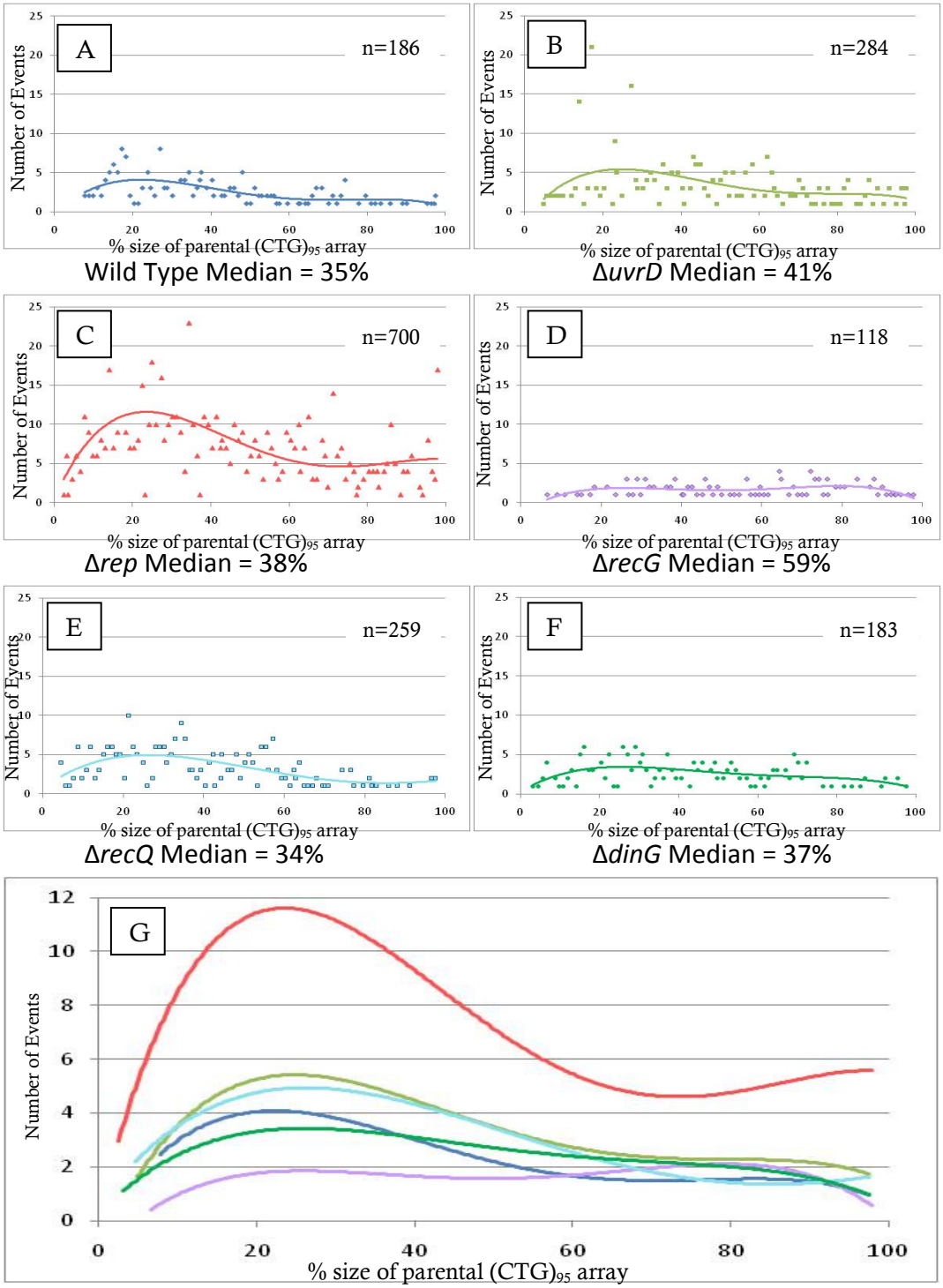

Figure 5.2: Distribution of deletion sizes of *E.coli* strains carrying the $(CTG)_{95}$ repeat array on the leading strand. Deletions are presented as a percentage of the 95 repeat array plotted against the frequency of the observed deletion size for strains: Wild Type (**A**), Δ*uvrD* (**B**), Δ*rep* (**C**), Δ*recG* (**D**), Δ*recQ* (**E**), Δ*dinG* (**F**). Trend lines for each distribution were plotted together (**G**).

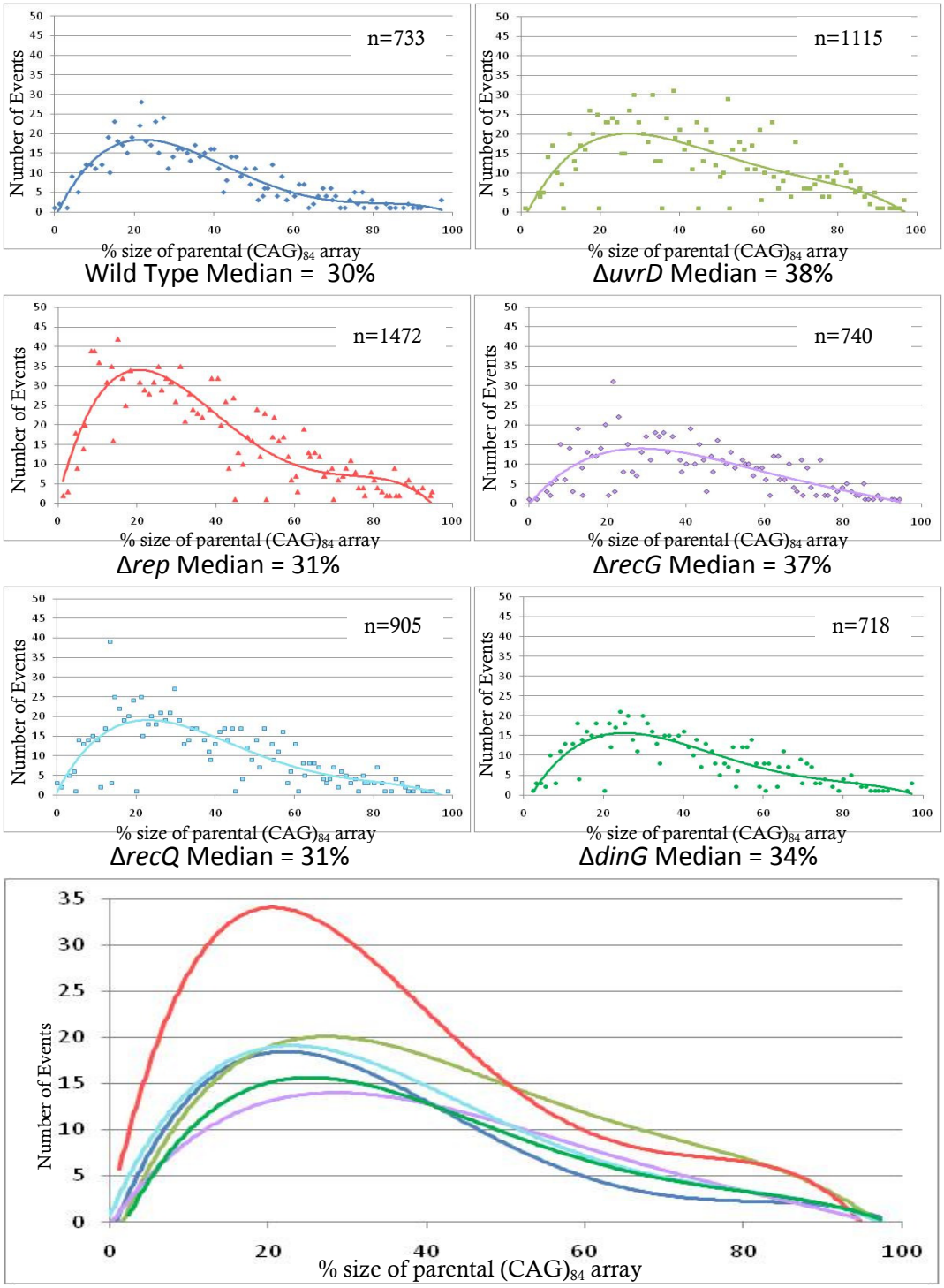

Figure 5.3: Distribution of deletion sizes of *E.coli* strains carrying the $(CAG)_{84}$ repeat array on the leading strand. Deletions are presented as a percentage of the 84 repeat array plotted against the frequency of the observed deletion size for strains: Wild Type (**A**), Δ*uvrD* (**B**), Δ*rep* (**C**), Δ*recG* (**D**), Δ*recQ* (**E**), Δ*dinG* (**F**). Trend lines for each distribution were plotted together (**G**).

5.2.2 Expansion instability in helicase mutants

5.2.2.1 Expansion instability proportion

Expansion instability proportions were calculated for each strain from 60 independent overnight cultures and plotted in Figure 5.4. Logistic regression analysis was carried out on the data to determine if any significant difference existed between any of the mutant strains tested and the wild type strain for a particular orientation of the repeat array (Table 5.2).

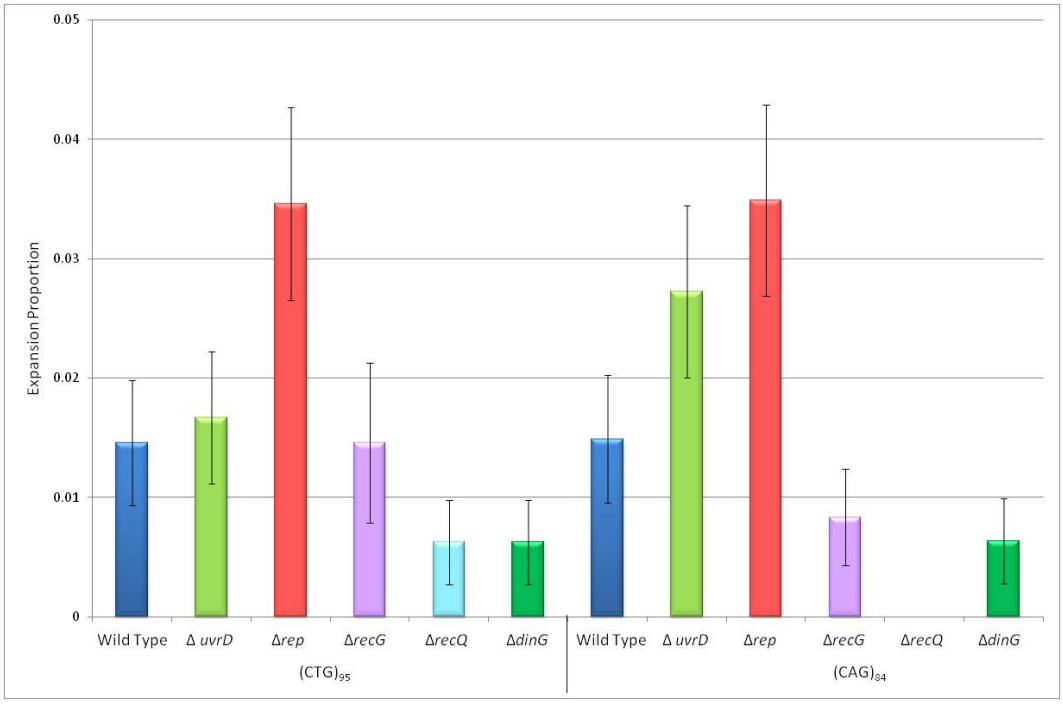

Figure 5.4: Expansion instability proportion of CTG•CAG arrays in *E. coli* strains containing a mutation in genes encoding helicase proteins. Each bar represents the proportion of expansion instability events calculated from 60 independent assays. Error bars show the standard error of the mean calculated for each strain.

Table 5.2: Logistic regression analysis of expansion instability proportions of helicase mutants. p≤ 0.05 considered statistically significant.

Orientation	Strain	P-Value
$(CTG)_{95}$	Δ*uvrD*	0.785
	Δ*rep*	0.001
	Δ*recG*	1.000
	Δ*recQ*	0.218
	Δ*dinG*	0.217
$(CAG)_{84}$	Δ*uvrD*	0.213
	Δ*rep*	0.063
	Δ*recG*	0.364
	Δ*recQ*	*
	Δ*dinG*	0.211

As observed in previous chapters many of the mutant strains tested did not show a dependence of expansion instability on repeat orientation relative to the origin of replication (Figure 5.4). In this study, the wild type, Δ*rep*, Δ*recG*, Δ*dinG* did not display an orientation dependence in expansion instability.

Even though the Δ*uvrD* mutant may appear to show an increased expansion proportion in the $(CAG)_{84}$ orientation compared to the wild type, logistic regression analysis shows that this is not a statistically significant difference.

Deletion instability proportion increased significantly in the $(CTG)_{95}$ Δ*rep* mutant but not in the equivalent strain in the $(CAG)_{84}$ orientation. Expansion proportion increased in this mutant strain compared to the wild type in both orientations ($p(CTG)_{95}$= 0.001; $p(CAG)_{84}$= 0.063 – close to the 0.05 statistical significance level). In both orientations, a deletion of the *rep* gene produces an

approximate 2-fold increase of the expansion instability proportion, indicating that the Rep protein acts in the wild type cell to prevent expansions.

Deletion of the *recQ* gene had no effect on the deletion instability of either orientation of the repeat array. Interestingly however, deletion of the *recQ* gene in $(CAG)_{84}$ strains suppressed the expansion instability altogether. In this orientation, repeat expansions are dependent on the RecQ protein, implicating an essential role of this protein in a wild type cell in the expansion instability pathway. In cells carrying TNR in the $(CTG)_{95}$ orientation, *recQ* deletion did not have such a pronounced effect. Expansion proportion was reduced in this orientation but not by a statistically significant amount ($p= 0.218$) suggesting that separate pathways of expansion may be acting in the different repeat orientations. Alternatively in a strain carrying (CAG) repeats on the leading strand template, deletion of the *recQ* gene may produce cell death in cells undergoing expansion events, suggesting RecQ protein may be essential for survival in cells undergoing TNR expansion mechanisms. This would be difficult to test as the frequency of expansion events is sufficiently low that no noticeable effect in viability may be seen.

Slightly reduced expansion proportions were also seen in the Δ*dinG* mutant strains carrying TNR in both orientations. Though they were not statistically significant, the low frequency of expansions in the wild type strains may obscure identification of genes having slight effects. To address this question the expansion instability was recalculated including expansions previously removed from the assay, as they were considered to have occurred

during the growth of the colony on the agar plate and not in the overnight culture (Figure 5.5).

This second analysis of the assay on expansion proportion confirmed many of the observations made using the first analysis method. There was no effect on expansion instability proportion of Δ*uvrD* or Δ*recQ* in the $(CTG)_{95}$ orientation, or of Δ*uvrD* or Δ*recG* in strains in the $(CAG)_{84}$ orientation. Statistical analysis (Table 5.3) almost suggested a potential role of the RecG protein in strains carrying a repeat array in the $(CTG)_{95}$ orientation. However, this result is not supported by the previous analysis and the potential increase in expansion proportion was relatively small suggesting it may not be a biologically significant difference (Figure 5.5).

Including expansion lengths produced during growth of a colony on the agar plate meant that the $(CAG)_{84}$ Δ*recQ* strain included a small number of expansion lengths (n= 3). Though there were only a small number and the effect of the mutation is still statistically significant (p= 0.021), it did mean that not all expansion instability was prevented by deletion of the *recQ* gene, indicating there must be an alternative source of expansion instability as well.

The Δ*dinG* mutant strains potentially showed a slight reduction in expansion instability proportion using the first analysis method. In this second analysis, expansion proportion of the $(CTG)_{95}$ Δ*dinG* strain was not reduced compared to the wild type strain. However, expansion instability of this mutant in the $(CAG)_{84}$ orientation appeared to be reduced compared to the wild type

strain. Even though this decrease in expansion instability was not to a significance of 95% (p= 0.091) it may still suggest a possible role of the protein in expansion pathways.

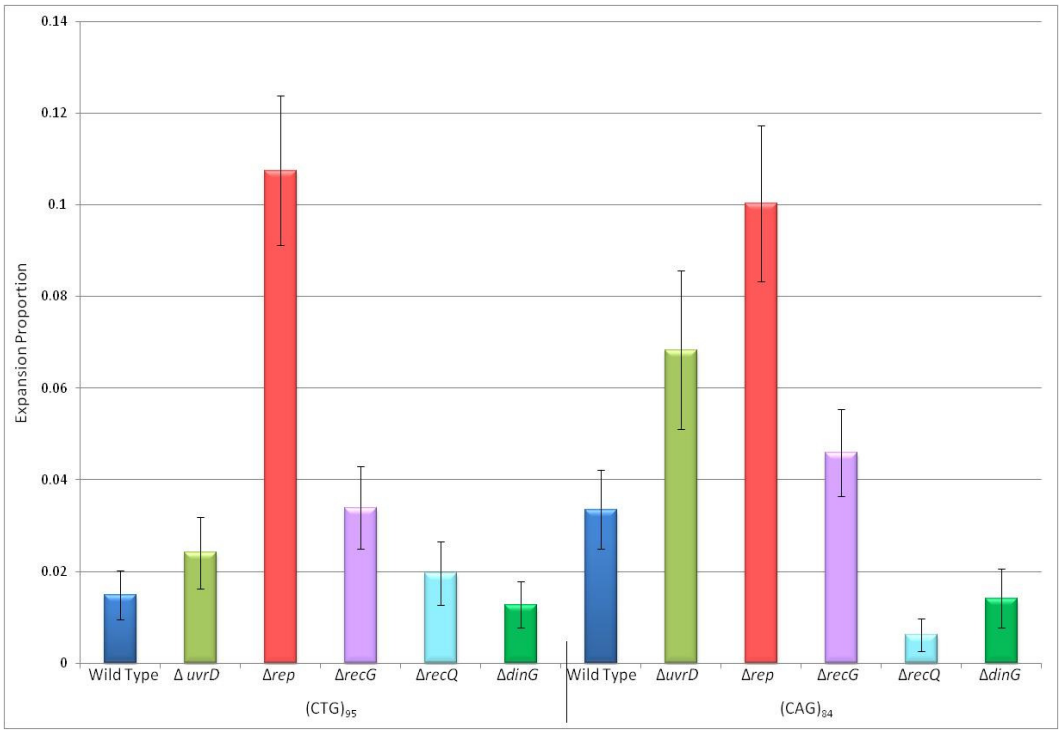

Figure 5.5: Expansion instability proportion of CTG•CAG arrays in *E. coli* cells containing a mutation in genes encoding helicase proteins, adjusted to include instability occurring during growth of colony on plate. Each bar represents the proportion of expansion events calculated from 60 independent assays including expansion events considered to have occurred during the growth of the colony on the agar plate. Error bars show the standard error of the mean calculated for each strain.

Table 5.3: Logistic regression analysis of expansion instability proportion in helicase mutants, adjusted to include instability occurring during growth of the colony on a plate. p≤ 0.05 considered statistically significant.

Orientation	Strain	P-Value
$(CTG)_{95}$	Δ*uvrD*	0.747
	Δ*rep*	<0.001
	Δ*recG*	0.061
	Δ*recQ*	0.615
	Δ*dinG*	0.783
$(CAG)_{84}$	Δ*uvrD*	0.072
	Δ*rep*	<0.001
	Δ*recG*	0.410
	Δ*recQ*	0.021
	Δ*dinG*	0.091

5.2.2.2 Expansion size

Expansion events are infrequent in *E. coli* cells, and do not form a normal distribution. Non-parametric analyses were carried out to determine if any of the mutant strains tested affected the size of expansion products observed. The Median expansion size was calculated for each strain and plotted (Figure 5.6) with error bars representing the inter-quartile range for the data. Kruskal-Wallis analysis was then conducted on the data to test for any significant differences between the strains tested (Table 5.4).

The median expansion lengths detected were all relatively small, with most not exceeding 110% of the parental TNR array size. No statistically significant difference on expansion size was found in any of the strains tested in either orientation of the repeat array. However, from the graph it appears that there may potentially be an effect of Δ*dinG* deletion in the strain carrying CTG

TNR on the leading strand template, which may be obscured by the large error bars present in the wild type strain for that orientation. This may suggest that the DinG protein is acting in a pathway to allow larger expansions in a cell carrying the gene.

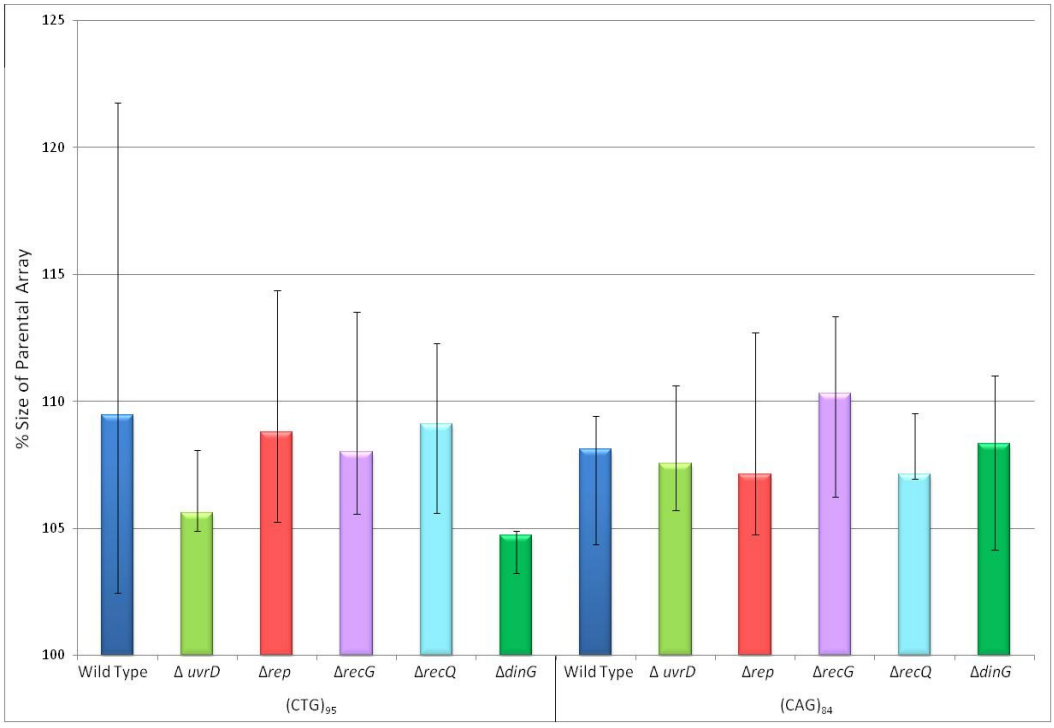

Figure 5.6: Median expansion size as a percentage of parental array length in *E. coli* helicase mutant strains carrying CTG•CAG repeat array. The median expansion size was calculated for each strain from each of the expanded lengths observed. Error bars represent the inter-quartile range for each median.

Table 5.4: Kruskal-Wallis analysis of expansion sizes in helicase mutants. p≤ 0.05 considered statistically significant.

Orientation	Strain	Median	Rank	P – value
$(CTG)_{95}$	Wild Type	109.5	47.7	
	Δ*uvrD*	105.6	39.9	
	Δ*rep*	108.8	54.0	
	Δ*recG*	108.0	50.4	
	Δ*recQ*	109.1	54.1	
	Δ*dinG*	104.7	29.8	
	Overall		49.5	0.328
$(CAG)_{84}$	Wild Type	108.1	47.1	
	Δ*uvrD*	107.5	51.7	
	Δ*rep*	107.1	61.3	
	Δ*recG*	110.3	51.4	
	Δ*recQ*	107.1	48.5	
	Δ*dinG*	108.3	53.7	
	Overall		52.5	0.827

5.2.3 Instability in replication restart mutants

The Rep protein was shown *in vitro* to have a role in the reinitiation of replication from stalled replication forks, in a pathway involving the protein PriC (Heller and Marians, 2005b). In order to test whether replication restart is a pathway involved in the instability of CTG•CAG TNR, strains carrying either orientation of the repeat arrays and a *priC* deletion were constructed. Instability assays were carried out on these strains and the results for both deletion and expansion instability were plotted (Figure 5.8 and Figure 5.9 respectively).

Deletion instability proportions were not significantly different from the wild type deletion proportion in Δ*priC* strains carrying either orientation of the repeat array (p= 0.322 and p= 0.225 for $(CTG)_{95}$ and $(CAG)_{84}$ respectively). In

strains in which the CTG repeats are present on the leading strand, deletion of the *rep* gene greatly destabilized the TNR array increasing deletion proportion approximately 6-fold. However, mutation of the *priC* gene in this orientation did not produce a similar effect, suggesting that PriC mediated replication restart is not responsible for the increased deletion instability seen in the *rep* mutant strain.

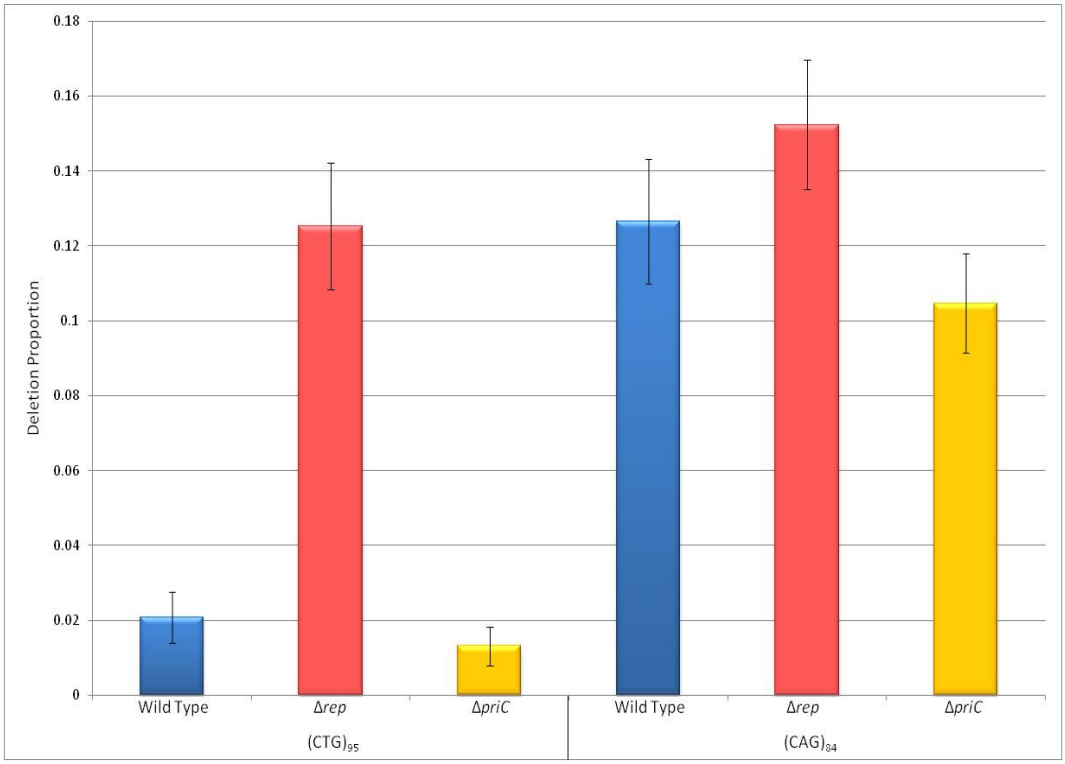

Figure 5.7: Deletion instability proportion of CTG•CAG arrays in *E. coli* strains containing a mutation in the *rep* or *priC* gene. Each bar represents the proportion of deletion instability events calculated from 60 independent assays. Error bars show the standard error of the mean calculated for each strain.

Table 5.5: logistic regression analysis of deletion instability in replication restart mutant strains. p≤ 0.05 considered statistically significant.

Orientation	Strain	P-Value
$(CTG)_{95}$	Δ*rep*	<0.001
	Δ*priC*	0.322
$(CAG)_{84}$	Δ*rep*	0.289
	Δ*priC*	0.225

Similarly, no significant effect on expansion instability was detected in *priC* mutant strains carrying TNR arrays in either orientation (p= 0.219 and p= 0.218 for $(CTG)_{95}$ and $(CAG)_{84}$ respectively). Neverthless, a mutation in the *rep* gene in strains carrying TNR arrays in either orientation produced an increase in expansion proportion. These data suggest the pathway involved in the action of Rep on TNR instability is not mediated by the PriC protein.

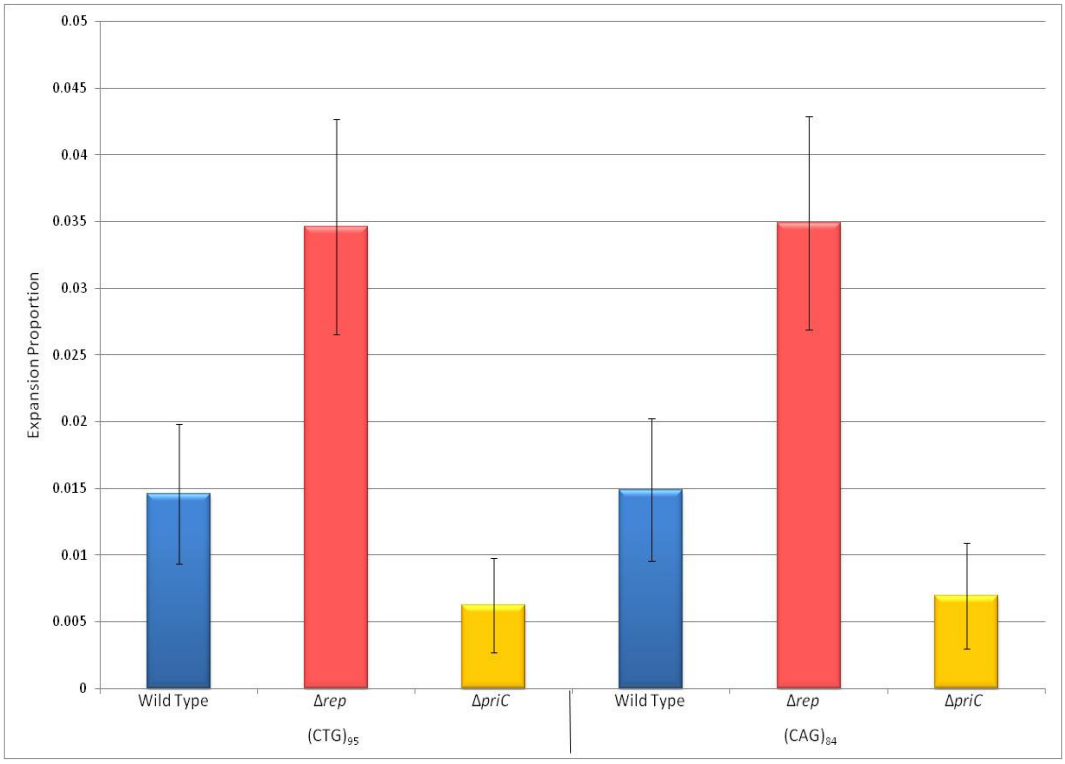

Figure 5.8: Expansion instability proportion of CTG•CAG arrays in *E.coli* strains containing a mutation in the *rep* or *priC* gene. Each bar represents the proportion of expansion instability events calculated from 60 independent assays. Error bars show the standard error of the mean calculated for each strain.

Table 5.6: logistic regression analysis of expansion instability in replication restart mutant strains. p≤ 0.05 considered statistically significant.

Orientation	Strain	P-Value
$(CTG)_{95}$	Δ*rep*	0.001
	Δ*priC*	0.219
$(CAG)_{84}$	Δ*rep*	0.063
	Δ*priC*	0.218

5.2.4 Instability in double mutants

Interestingly, expansion instability data presented in this Chapter suggested that the Rep protein acts in an otherwise wild type cell to prevent expansion instability, while the RecQ protein is involved in the pathway causing expansion in strains carrying the CAG TNR array on the leading strand. There was potentially a role of DinG protein in generating expansion instability in strains carrying the CAG repeat array on the leading strand. To answer the question of whether the expansions detected in a *rep* mutant strain were dependent on the RecQ or DinG proteins, double mutant strains of Δ*recQ*Δ*rep* and Δ*dinG*Δ*rep* were constructed. The TNR instability in these strains was tested and compared to the instability of the single mutants (Figure 5.9).

It was predicted that the increase in expansions seen in the Δ*rep* strain carrying CAG repeats in the leading strand, would be dependent on RecQ, so the double mutant was predicted to have a significantly lower expansion proportion then the *rep* mutant. The Δ*dinG*Δ*rep* double mutant was also predicted to have a slight effect on expansion proportion in the CAG leading strand orientation, compared to the Δ*rep* single mutant. Single mutants in *recQ* or *dinG* did not produce a significant effect in the CTG leading strand orientation so no effect of the double mutation in this orientation was expected.

Figure 5.9 shows that in Δ*recQ*Δ*rep* and Δ*dinG*Δ*rep* double mutant strains carrying the $(CTG)_{95}$ repeat array on the leading strand the level of expansion instability was not significantly different than in the Δ*rep* single mutant. Both

double mutants display an expansion proportion very similar to that of the Δ*rep* single mutant. This was to be predicted as the single mutants did not have a significant effect on strains carrying repeats in this orientation either.

Deletion of the *recQ* gene prevented expansion instability in an otherwise wild type cell carrying the CAG array on the leading strand template. Surprisingly, the Δ*recQ*Δ*rep* double mutant strain presented a level of expansion instability proportion close to that of the Δ*rep* single mutant strain. This result proves that in a Δ*rep* mutant strain, unlike in a wild type strain, expansion instability is not dependent on RecQ protein, suggesting a different pathway of expansion instability may be involved.

Deletion of both the *dinG* and *rep* genes in the orientation in which $(CAG)_{84}$ is present on the leading strand did however have a significant effect. The expansion proportion detected in this strain was approximately 3-fold lower than that of the Δ*rep* single mutant in this orientation. The decrease of expansion instability due to the absence of DinG is dramatically more pronounced in a Δ*rep* mutant than in a wild type background. Table 5.7 shows the results obtained when a logistic regression model was applied to determine whether the Δ*dinG*Δ*rep* double mutant was significantly different to either of the single mutant controls. While the expansion instability is significantly different to the Δ*rep* mutant strain, it is not different to the Δ*dinG* mutant, confirming that in a Δ*rep* mutant background expansion instability is partially dependent on DinG protein.

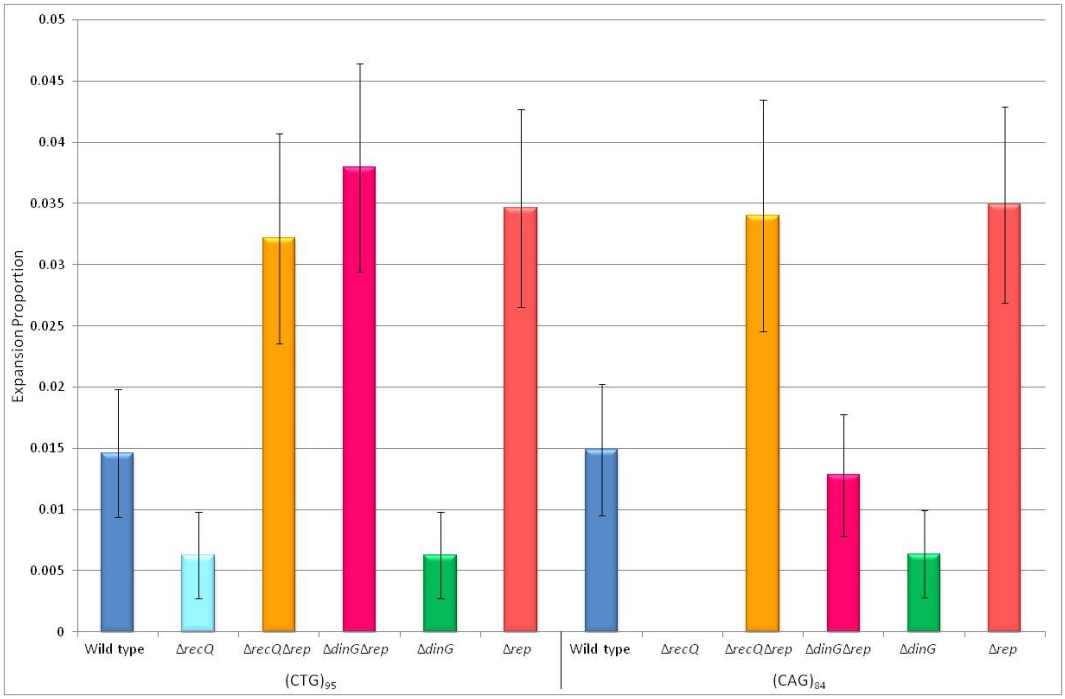

Figure 5.9: Expansion instability proportion of CTG•CAG arrays in *E.coli* strains containing a single or double mutation in genes encoding helicase proteins. Each bar represents the proportion of expansion instability events calculated form 60 independent assays. Error bars show the standard error of the mean calculated for each strain.

Table 5.7: Logistic regression analysis of expansion instability of single *rep* and *dinG* helicase mutants compared to the double mutant strain. p≤ 0.05 considered statistically significant.

Orientation	Strain	P-Value
$(CAG)_{84}$	Δ*rep*	0.036
	Δ*dinG*	0.320

Analysis of the size of expansion products obtained in the double mutant strains studied was carried out by calculating the median expansion size for each strain;

these data were plotted in Figure 5.11. The graph shows that in the double mutants studied, both the Δ*recQ*Δ*rep* and the Δ*dinG*Δ*rep* strains, there is no significant difference in the size of expansions compared to the single mutants or the wild type strains.

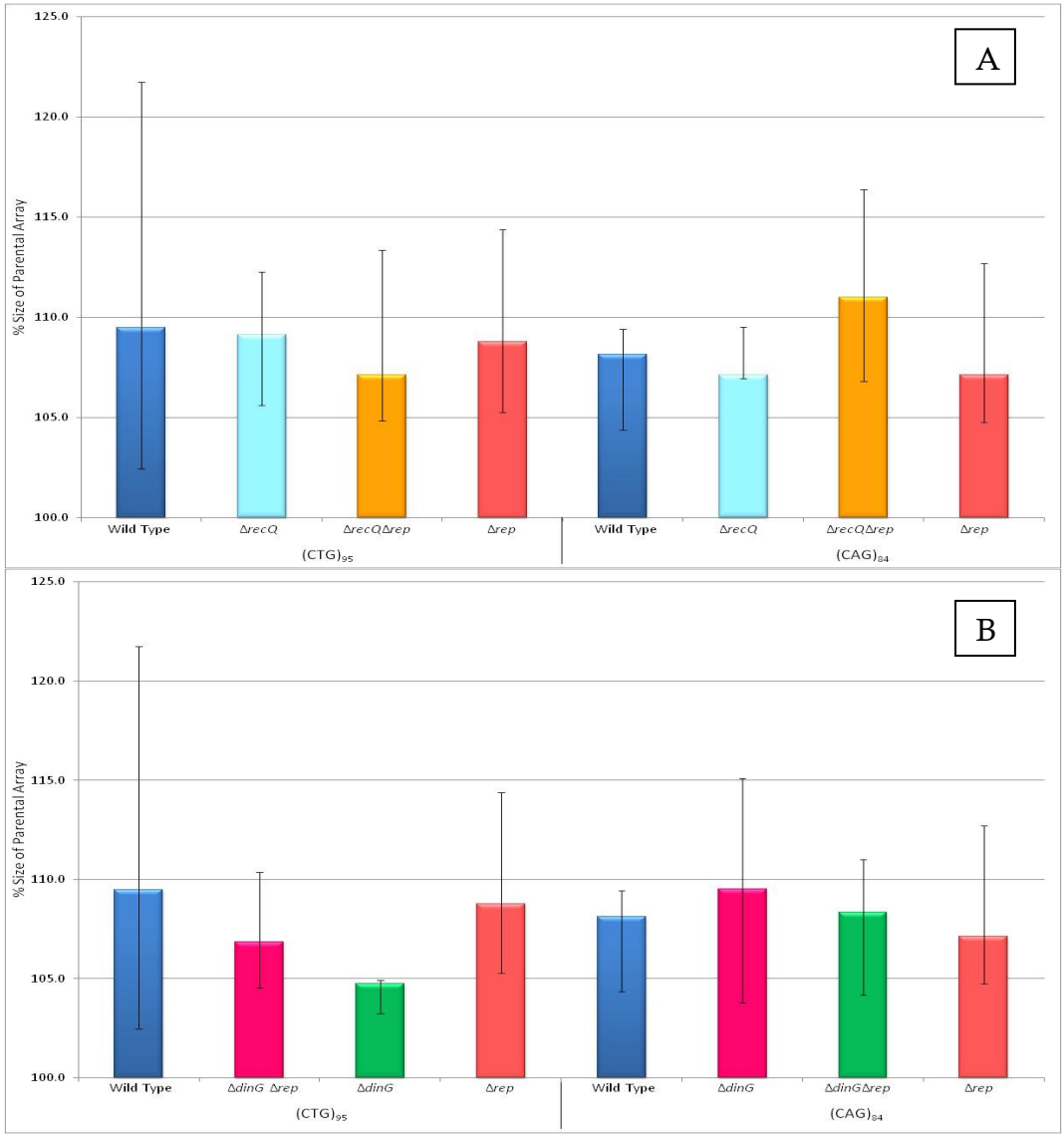

Figure 5.10: Median expansion size as a percentage of parental array length in *E. coli* strains carrying CTG•CAG repeat array. Panels show the median expansion size in the Δ*recQ*Δ*rep* mutants (**A**) and Δ*dinG*Δ*rep* mutants (**B**). The median expansion size was calculated for each strain from each of the expanded lengths observed. Error bars represent the inter-quartile range for each median.

5.3 Discussion

5.3.1 Expansion instability is orientation independent

The data presented in this chapter showed that deletion instability of TNR arrays is dependent on the orientation of the array relative to the origin of replication, for all but the Δ*rep* strain. Orientation dependence of deletion instability is in agreement with the other results presented in previous chapters and with published work on the subject (Kang *et al.*, 1995a). Interestingly for most of the strains studied, expansion instability is not dependent on the orientation of the TNR array relative to the origin of replication. These data suggest that either the ability of sequence to form into thermodynamically stable secondary structures is not the limiting factor in the instability mechanism, or that structure formation is occurring on both nascent leading and lagging strands of the replication fork.

5.3.2 UvrD helicase reduces deletion instability

Strains in which the *uvrD* gene was deleted displayed increased deletion instability in strains carrying TNR arrays in either orientation. It was established in Chapter 3 that this effect was not due to the role of UvrD helicase in the replication fork reversal pathway, and it was hypothesized that it may be due to its helicase activity. The helicase may load onto the replication fork in the transiently single stranded region of the lagging strand template. UvrD helicase translocates in a 3'→5' direction so would move away from the apex of the fork

along the lagging strand template. If a TNR array was in this region and had folded into either a looped out or a hairpin-like structure, depending on the sequence, the helicase would be able to separate the intrastrand bonds holding such a structure together, thereby 'ironing' it out. In UvrD^{+} cells this would prevent deletion instability, while in cells lacking the protein such structures would persist, leading to deletion in a subsequent round of replication (Figure 5.12).

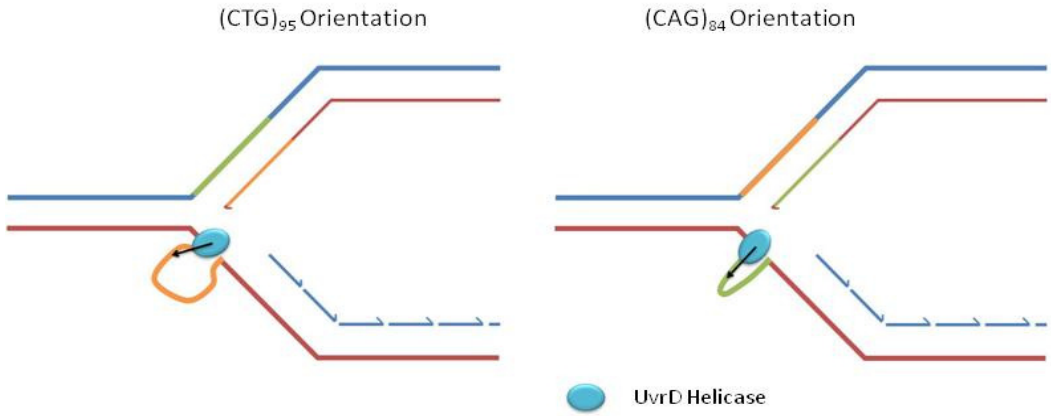

Figure 5.12: Model of UvrD helicase action on CTG•CAG TNR arrays in the *E. coli* chromosome. The helicase can unwind structures, looped out DNA or hairpin-like structures, present in the transiently single stranded region of the lagging strand template.

Deletion of the *uvrD* gene had no significant effect on expansion instability of CTG•CAG TNR tracks. If expansion instability is generated as a result of structures forming on the nascent strands, it could be that the UvrD helicase is unable to access these structures, preventing it from 'ironing' them out. *In vitro* experiments on the UvrD protein and its *S. cerevisiae* homologue Srs2 demonstrated that Srs2 was able to unwind hairpin-like structures formed by TNR at a 3-fold quicker rate then UvrD helicase (Bhattacharyya and Lahue,

2005). This corresponds with srs2Δ yeast cells displaying increased expansion instability compared to the wild type cell (Bhattacharyya and Lahue, 2004) unlike results presented here for Δ*uvrD E. coli* strains.

5.3.3 Deletion instability is increased in a *rep* mutant strain

Interestingly, analysis of the *rep* mutant strain did not show the orientation dependence of CTG•CAG deletion instability normally detected. Deletion of the gene encoding this helicase in strains carrying a $(CTG)_{95}$ TNR array on the leading strand produced an increase in deletion proportion of approximately 6-fold. On the other hand, in strains carrying TNR in the opposite orientation, mutation in the *rep* gene had no significant effect on the deletion instability proportion observed.

To investigate this result further, the instability of a *priC* mutant was also studied. The Rep and PriC proteins work together in a pathway that allows re-initiation of replication from arrested replication forks (Heller and Marians, 2005b). No effect of *priC* mutation was seen on the deletion instability of strains carrying TNR arrays in either orientation, ruling out the replication restart pathway from being involved in the deletion instability seen in a *rep* mutant.

Additionally, Rep protein has been shown to associate with the DnaB protein at the replication fork (Guy *et al.*, 2009), and it is able to remove protein bound to the DNA ahead of the oncoming replication fork (Boubakri *et al.*, 2010), which may explain why time taken to complete a replication cycle is

longer in a *rep* mutant cell (Lane and Denhardt, 1975). Here, the model I propose to explain the deletion instability effect of a *rep* mutant strain, depends on the Rep protein's ability to move with the replication fork, clearing proteins bound to DNA from ahead of it (Figure 5.13). CTG repeat arrays form much more thermodynamically stable structures than CAG repeat sequences (Gacy and McMurray, 1998). As lagging strand synthesis is discontinuous, when a CTG repeat array is present in the single stranded region of the lagging strand template it can form into a hairpin structure, which following a subsequent round of replication would produce a deletion of the repeat array. CAG repeats in similar circumstances would form a much less thermodynamically stable looped-out structure, less likely to survive into a second round of replication. A strain mutated in the *rep* gene may experience more replication fork arrest as the helicase is not present to remove DNA bound proteins ahead of the fork. In this case it may be that the nascent leading strand at the fork is unwound by the action of a helicase protein generating single stranded DNA which would be a substrate for degradation by Exonuclease I (a 3'→5' nuclease). This mechanism would reveal the CTG array on the leading strand template, allowing it to fold into a thermodynamically stable hairpin structure (Figure 5.12). Mutation of the *rep* gene in strains carrying the CTG array on the lagging strand template would have no effect as the sequence on the leading strand template is composed of CAG TNRs which form less thermodynamically stable structures. The single stranded DNA present on the lagging strand template would still be the site of the majority of events leading to deletion instability.

If instability events in a *rep* mutant strain are due to replication fork stalling in the proposed manner, alternative methods of fork stalling, such as hydroxyurea treatment or studying instability in a *dnaB*ts mutant, could be employed to test this model. Hydroxyurea treatment depletes deoxyribonucleic acid pools in the cell which leads to stalling of the replication fork. DnaB is the replicative helicase, and though it is essential temperature sensitive mutants exist in which the helicase can be inactivated by shifting to the non-permissive temperature. Using either of these systems would produce a stall similar to that proposed to be occurring in the Δ*rep* mutants studied here.

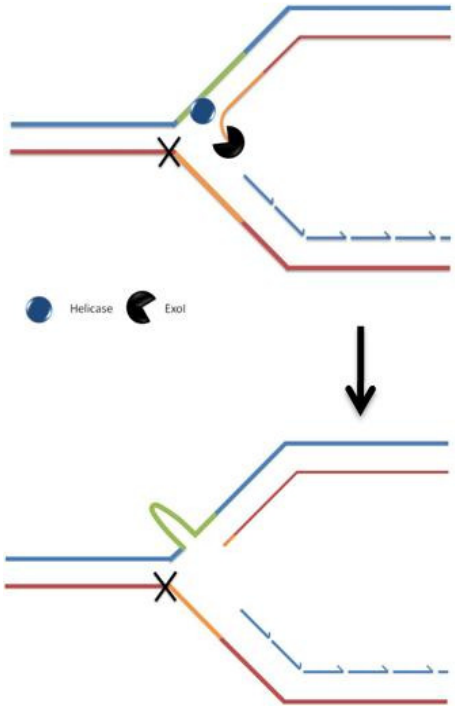

Figure 5.13: Model of deletion instability intermediates produced in Δ*rep E. coli* strains carrying CTG TNR array. *rep* gene mutation leads to fork stalling, which could allow helicase and exonuclease action to unwind and the degrade the nascent leading strand, revealing the CTG TNR array present on the template strand. This could form into a hairpin structure which is an intermediate in the deletion instability pathway.

5.3.4 Mutation of the *recQ* gene reduces expansion instability in strains with a CAG array on the leading strand template

Deletion of the gene encoding the RecQ protein had no effect on the deletion instability proportion detected compared to the wild type, in strains carrying TNR arrays in either orientation. Similarly, the mutation had no effect on expansion instability in strains carrying a CTG array on the leading strand template. However, when the TNR array was in the opposite orientation relative to the origin of replication, mutation of the *recQ* gene severely restricted expansion instability. This result suggests that in a wild type cell carrying a CAG array on the leading strand template, RecQ protein is facilitating expansion events. Or alternatively deletion of the *recQ* gene is lethal in cells undergoing expansion events.

Expansion events are believed to be the result of structure formation in the nascent strands, which following a further round of replication would produce an expanded TNR array. RecQ is a 3'→5' helicase capable of binding to numerous different substrates. A model has been developed (Figure 5.14) to explain the presented data. A progressing replication fork may stochastically encounter a challenge that causes it to stall on the leading strand, separating leading and lagging strand synthesis. In the event of such a stall RecQ protein may bind to the nascent leading strand and translocate in a 3'→5' direction, dissociating it from its template. Due to RecQ action, the TNR present on the nascent strand could fold into a secondary structure. Following repair of this fork arrest and reinitiation of replication if the secondary structure formed on the

nascent leading strand were to persist it would lead to an expansion event. The strain in which *recQ* deletion was shown to have an effect carries a CTG array on the nascent leading strand, which would produce a thermodynamically stable hairpin-like structure in such circumstances. Helicases have been shown to be able to remove structure in DNA, for example the Srs2 protein in *S. cerevisiae* (Bhattacharyya and Lahue, 2004). However, this model requires helicase action to facilitate the formation of structured DNA. It has been shown that there are helicase proteins that are able to unwind DNA and allow re-annealing in such a way that structures form. For example, though not implicated in the formation of hairpin structures, both the *E. coli* RuvB helicase and the *S. cerevisiae* Rad5 helicases have been shown to facilitate formation of 4-way DNA structures (Blastyak *et al.*, 2007; Seigneur *et al.*, 1998).

When a replication fork arrest leading to a gap on the leading strand template, this single stranded DNA would be coated by SSB protein. RecQ physically interacts with SSB protein, and the helicase activity of RecQ is stimulated by this interaction (Shereda *et al.*, 2007). These data could suggest that RecQ protein is being recruited to sites of leading strand gaps by the SSB protein. The human RecQ family member, WRN protein, has been shown to act on substrates in which an arrested fork presents a gap on the leading strand (Machwe *et al.*, 2007). The WRN protein has both helicase and 3'→5' exonuclease activity, the latter was shown to be useful in degrading a small amount of the nascent leading strand at an arrested fork, enhancing the unwinding of the template/daughter duplex. *E. coli* RecQ protein does not

possess exonuclease activity, but it is conceivable that another protein could substitute for this role in the bacterium. ExoI, a 3'→5' *E. coli* nuclease, would be a suitable candidate, as it has been shown to associate with the RecQ protein, probably through SSB interaction (Shereda *et al.*, 2007).

Additionally, another helicase might be required to bind the template strand, separating a free 3'-end to which RecQ can bind. However, this protein has yet to be identified.

When expansion events considered to have occurred during the growth of the colony on an agar plate was included in the analysis, some expansion instability was detected in the $(CAG)_{84}$ Δ*recQ* strain. This result suggests that while the proposed model is a significant pathway of TNR expansion in *E. coli* it may not be the only one.

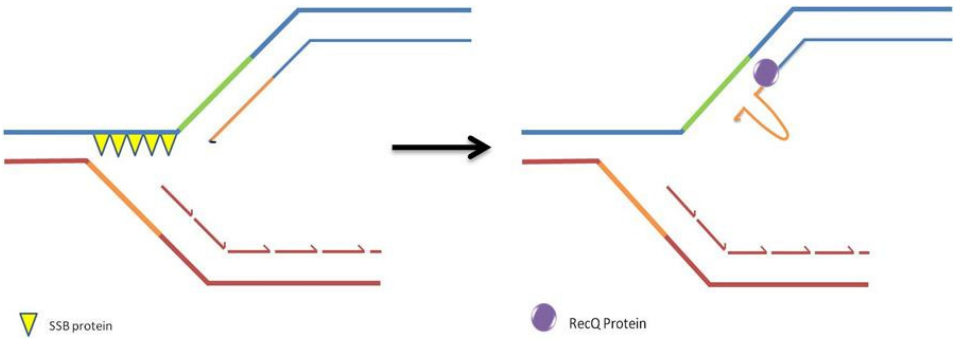

Figure 5.14: Model of RecQ dependent expansions in *E. coli* strains carrying a CAG TNR array on the leading strand template. A stall to replication could cause the leading strand synthesis to stall, producing a leading strand gap. RecQ helicase action on the nascent leading strand could allow structure formation leading to expansions.

Importantly, expansion instability levels were similar in strains containing either orientation of the CTG•CAG array, but this level was only dependent on the presence of RecQ protein when the CAG array was on the leading strand template. Therefore, strains carrying a TNR array in the opposite orientation relative to the origin of replication, could be exposed to another mechanism expansion instability. One could predict a model in which a helicase could be unwinding the nascent lagging strand allowing structure to form within. The helicase responsible for this however remains to be identified. Alternatively a model similar to that seen in *Fen1/Rad27* mutant cells in *S. cerevisiae* could be occurring, in which expansion instability occurs in this orientation due to structure formation on the nascent lagging strand during the flap processing of Okazaki fragments (Liu and Bambara, 2003).

5.3.5 Expansion instability increases in *rep* mutants carrying TNR arrays in either orientation

While the effect of a mutation in the *rep* gene was limited to the $(CTG)_{95}$ orientation for deletion instability, there was a clear effect on expansion instability in both orientations of the TNR array. Expansion proportion increased approximately 2.5-fold in strains carrying CTG•CAG repeats in either orientation.

Subsequent analysis of strains mutated in the replication restart gene *priC* showed that PriC had no effect on the expansion instability observed. PriC and

Rep helicase have been shown *in vitro* to work together in one pathway of re-initiation of replication of arrested forks (Heller and Marians, 2005b). The observation that a *priC* mutation has no effect on expansion instability, suggests that it is not the role of Rep helicase in this pathway that is involved in expansion instability.

A model was developed to account for the increase in expansion instability observed in *rep* mutant strains (Figure 5.15). This model is based on the role of Rep helicase in associating with DnaB protein (Guy *et al.*, 2009) and clearing DNA bound proteins ahead of the replication fork (Boubakri *et al.*, 2010). In strains deficient for the Rep helicase protein, DNA bound proteins may cause arrest of the replication fork. When TNR arrays are present at the site of such a stall, unwinding of the nascent and template strand by helicase proteins could allow a structure formation. As no orientation dependence on expansions was observed this structure formation is predicted to be occurring at either the nascent leading or lagging strand with equal likelihood. In the CTG orientation (Figure 5.15a), a helicase could unwind the nascent lagging strand from its template, allowing a structure to form in the now single stranded CTG array on that strand. Strains carrying TNRs in the CAG orientation are predicted to expand by helicase action of unwinding the nascent leading strand, thus allowing structure to form (Figure 5.15b).

As the data showed most expansions in strains carrying the TNRs in the CAG orientation were dependent on RecQ protein. Therefore, it was hypothesized that RecQ could be the helicase acting in the expansion pathway of

strains in the CAG orientation. However, in a Δ*recQ*Δ*rep* double mutant in this orientation, expansion instability was not dependent on the RecQ protein, indicating that the expansion events detected in the Δ*rep* mutant were not RecQ dependent as those in wild type cells are. Alternatively it could be the case that if mutation of the *recQ* gene leads to lethality in cells undergoing an expansion event, thus death is dependent on the presence of Rep helicase. In the Δ*rep*Δ*recQ* strain expansions would occur and the cells survive.

Work on the expansions of TNR arrays in *recQ* and *rep* mutant strains shows the phenotypes to be epistatic. It is only in strains expressing Rep protein that the effect of *recQ* deletion is seen. The model of instability above proposed to occur in a Δ*recQ* mutant cell (Figure 5.14) therefore requires the presence of Rep helicase.

The instability assay results suggested there may be some role of the DinG helicase in expansion instability, which was confirmed when double Δ*dinG*Δrep mutants were investigated. No effect of *dinG* mutation was seen on the observed expansion instability in strains carrying CTG TNR on the leading strand; but when the CAG TNR array was present on the leading strand, orientation expansion proportion in the Δ*dinG*Δ*rep* double mutant more closely resembled results obtained for the Δ*dinG* single mutant than the Δ*rep.* These data suggest that the expansion events occurring in the Δ*rep* strains are dependent, to some degree, on the presence of the DinG helicase.

The model developed proposes that in a strain lacking functional Rep helicase a replication fork can arrest, as the replicative helicase DnaB is unable to unwind the parental duplex ahead of the fork due to a persistent protein bound to the DNA. In such arrest sites, DinG helicase may bind to the leading strand of the replication fork, unwinding the template and newly synthesized strands. This action would generate single stranded DNA, and as the newly synthesized strand contains the CTG TNR array in this orientation, it can fold into a hairpin-like structure, which would lead to an expansion event following a subsequent round of replication. *In vitro* studies into the substrate specificity of the DinG protein found that a 5'-tailed substrate of 11-15nt was sufficient for protein binding (Voloshin and Camerini-Otero, 2007), suggesting that only a small leading strand gap would be required in this model.

The fact that a Δ*dinG*Δ*rep* double mutant still retains some degree of expansion instability may be indicative of the redundancy within helicase proteins, suggesting there may be another protein able to generate expansions in a *rep* mutant cell.

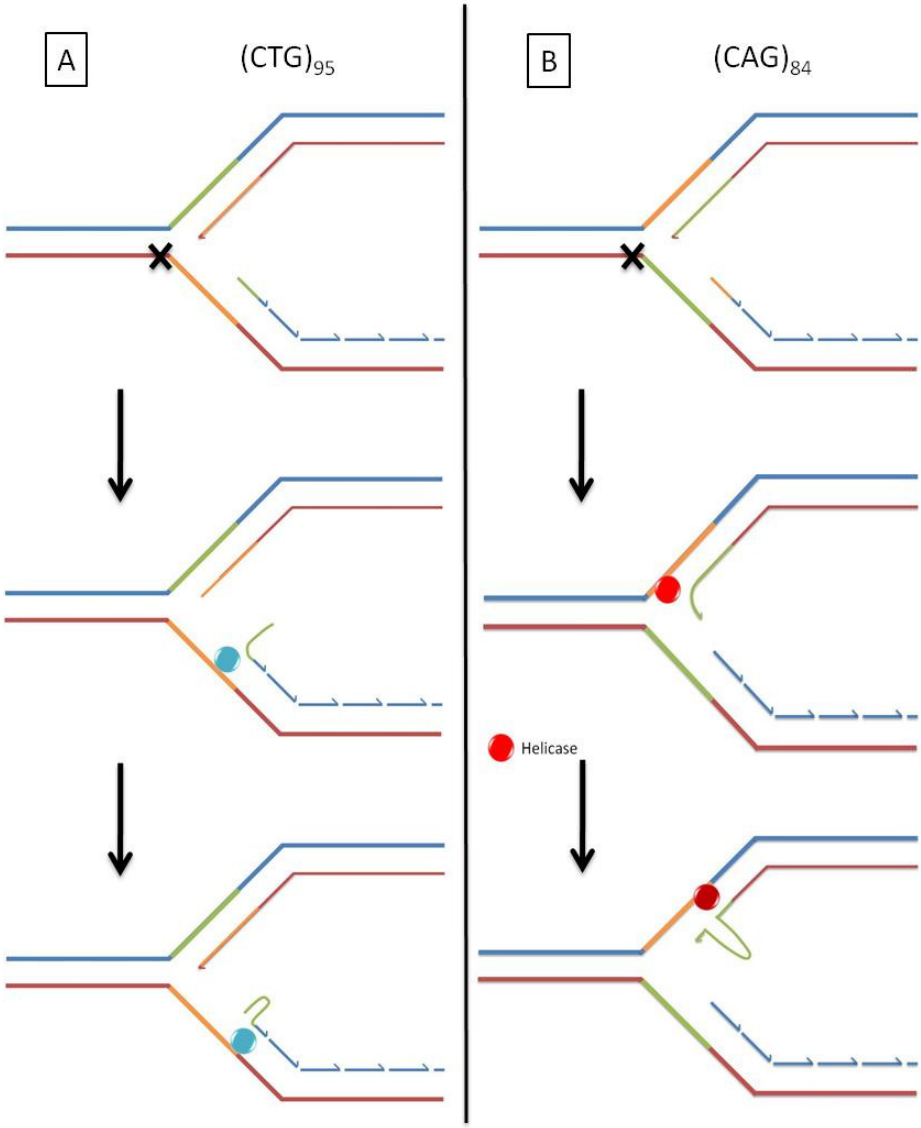

Figure 5.15: Model of expansion instability generated in *E. coli* Δ*rep* strains carrying CTG•CAG TNR arrays. **A**. Instability in strains carrying TNR in the CTG orientation may result from unwinding of the nascent lagging strand from its template by a helicase moving along the template strand, though as yet no evidence of this has been found. **B**. In the CAG orientation nascent leading strand may be unwound by DinG helicase allowing structure to form in the CTG TNR present.

5.3.6 RecG protein promotes larger deletions in strains carrying the CTG array on the leading strand template.

Analysis of the size of deletion products obtained by the strains studied showed that in most strains instability events favour large deletions, producing a negative skew in the deletion distribution. However, strains carrying the CTG repeats on the leading strand template that are mutated in the *recG* gene do not show this same negative skew, the deletion distribution curve produced for this strain is much more even. Thus, the RecG protein is affecting the size of the deletion product obtained while not affecting the number of deletion events taking place. Strains carrying TNR in the opposite orientation did not show this effect of *recG* mutation.

This effect of *recG* mutation on deletion size, is similar to that noticed in the previous chapter for *recF*, *recO* and *recA* mutants, though more pronounced. In the previous chapter mutation of *recFOA* altered the deletion distribution so there was a slightly less negative skew, while in the *recG* mutant all bias is lost. This is similar to results published previously by the laboratory for strains carrying a TNR array in the same CTG orientation and mutated in the *sbcDC* genes (Zahra *et al.*, 2007). The SbcCD protein complex in *E. coli* is related to the Rad50/Mre11 complex in both humans and yeast (Sharples and Leach, 1995), both of which are involved in repair of double stranded breaks in DNA (Johzuka and Ogawa, 1995). One of the functions of the SbcCD complex is as an dsDNA exonuclease, in which it can act on a variety of DNA substrates including DNA hairpins (Bzymek and Lovett, 2001; Connelly *et al.*, 1999).

Figure 5.16 shows the proposed model to explain the *recFO*, *recA* and *recG* mutant results obtained in this work, and the *sbcDC* mutant strains studied previously in the lab (Zahra *et al.*, 2007). A CAG TNR array on the lagging strand template may form a looped out DNA structure when transiently single stranded (Figure 5.16a). Zahra and colleagues (Zahra *et al.*, 2007) showed that this species of intermediate was subject to the proofreading activity of DnaQ, which could prevent deletion instability arising from persistence of the structure. If the structure avoids proofreading by DnaQ, a slippage event could allow the newly synthesized DNA to anneal either side of the looped out structure, with an equilibrium being established between the two species (Figure 5.16b). CAG TNR arrays form less thermodynamically stable structures than CTG repeats. However, if the looped out structure were closed off by a slippage event (such as in Figure 5.16b) the ability of the loop out to flatten would be lost, and it may be favourable for the sequence to form a hairpin-like structure (Figure 5.16c). Such hairpin like structures might even be stabilized by the presence of proteins such as MutS, which could bind to the structure allowing it to maintain its form (Kovtun and McMurray, 2001).

DNA synthesis of the nascent lagging strand would trap such species irreversibly in these structures (Figure 5.16d). The Exonuclease SbcCD is able to recognize and cleave hairpin, and if present here, could cleave the CAG hairpin (Figure 5.16e) leaving a gapped structure. This gapped structure could be repaired by RecFOR mediated gap repair, which would form Holliday junctions (Figure 5.16f) which may be dependent on RecG for resolution (Figure 5.16g).

CAG structures which formed but were inaccessible to SbcCD would produce deletion products of different lengths, with no bias towards big deletions. As SbcCD recognizes large hairpins some of the bigger structures may be susceptible to cleavage by the exonuclease, biasing towards large deletions. Those structures attacked by SbcCD lacking one of the *recFOR* or *recA* genes would not enter the gap repair pathway and so may be repaired by another pathway in which some of the bias towards big deletions is lost. Work in chapter 4 showed that strains in the CTG orientation, carrying CAG TNRs on the lagging strand template, mutated in one of the *recFOR* or *recA* genes displayed less of a negative skew than a wild type strain carrying a TNR array in the same orientation. Some negative bias was still observed, consistent with the model that hairpin cleavage by SbcCD has occurred. Similarly strains mutated in *recG* show a different deletion distribution, in this case however, all bias is lost. If the model in Figure 5.16 were followed cell lacking RecG protein might be trapped in the species shown in Figure 5.16f. This problem could cause those cells which have entered this SbcCD pathway to die as they are unable to resolve the Holliday junctions formed. In such cases deletion distribution would resemble those cells lacking SbcCD as they would be the only ones to survive, such cells would not be biased towards big deletions and would show a flat deletion distribution, as seen for the $(CTG)_{95}$ Δ*recG* strain.

This model could be tested by construction of a Δ*recF*Δ*recG* double mutant. As the model predicts that RecF acts in the pathway before RecG, the

double mutant should not display the RecG phenotype as it would not proceed to structure F.

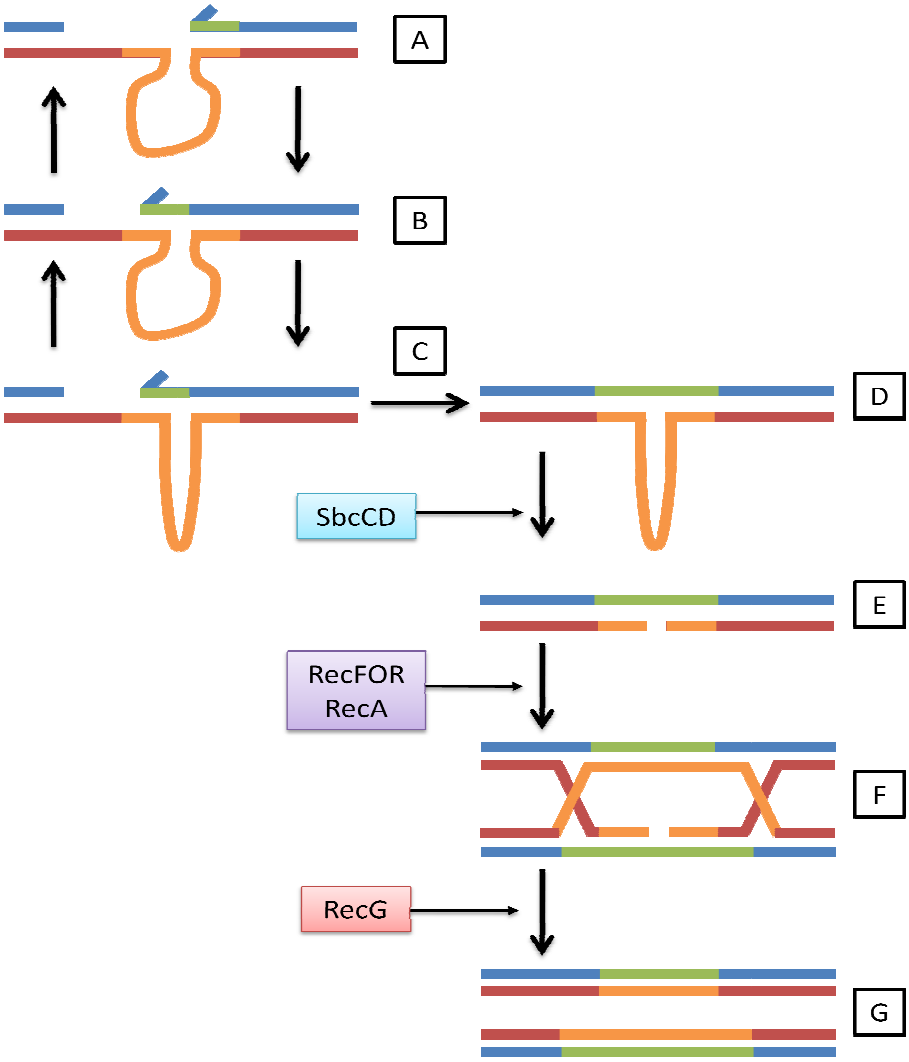

Figure 5.16. Model of deletion pathway in strains carrying CAG TNR on the lagging strand template. Looped out DNA formed in CAG TNR arrays may be stabilized by a slippage of the complementary strand, in which case they may form a hairpin structure. Hairpins are substrate for the nuclease SbcCD. The action of which could produce a gap requiring repair by RecFOR, RecA and RecG. This pathway could produce a bias towards large deletions.

Work in this chapter has shown that helicase action plays an important role in the instability of CTG•CAG TNR arrays. Deletion instability was affected by both Rep (in one orientation) and UvrD helicase (in both orientations). While the size of deletion products was affected by the RecG protein. Expansion instability was found to be orientation independent and it was proposed that this was due to the ability of stable structure to form on either the nascent leading or lagging strands with equal likelihood. This result is not entirely surprising as the nature of the nascent strand in a replication fork requires dissociation from its template in some way before structure formation can be achieved, differing from deletion instability in which there is always a region of template ssDNA present at the replication fork. The nature of arrested fork structures lead to different pathways of expansion instability, with those in a *rep* mutant cell differing from those in a wild type strain. However, the role of helicase proteins was key to the instability in both pathways.

Chapter Six:

Concluding Remarks

6.1 Summary of work presented

The work in this thesis presents an investigation of factors affecting deletion and expansion instability in chromosomally located CTG•CAG TNR arrays in *E. coli*. Previous studies of TNR instability in the bacterium have utilized plasmid based studies which present their own limitations. Using a system setup in this laboratory by a previous PhD student, Rabaab Zahra, this research avoids the complications of the plasmid based systems by studying CTG•CAG arrays that have been inserted into the chromosomal *lacZ* gene (Zahra *et al.*, 2007). I believe the most meaningful results discovered in the process of this work are, firstly, that expansion instability, unlike deletion instability, is independent of the orientation of the TNR array relative to the origin of replication. Also of interest was the observation that expansions were dependent on RecQ helicase, in the orientation in which CAG repeats were present on the leading strand.

Replication fork reversal was initially an interesting model for instability of TNR arrays as previous work had suggested a model for expansion instability involving this process (Mirkin, 2007). This model was tested by mutation of the genes *uvrD* and *recF*. UvrD is essential for RFR in certain replication mutant backgrounds, as it can clear toxic RecA filaments from ssDNA from an arrested replication fork. In cells carrying a mutation in the gene *recF*, RecA protein is not loaded onto its ssDNA substrate, thus negating the need for UvrD protein (Flores *et al.*, 2005). No significant role of the UvrD mediated pathway of RFR was detected on expansion instability. However, mutation of the *uvrD* gene

affected the deletion instability of CTG•CAG TNR arrays, suggesting a role for the protein in stabilizing them in *uvrD*$^{+}$ cells.

No significant effect of mutation of the genes *recF, recO,* or *recA* was found on the frequency of expansion or deletion events in the mutant strains tested. However, strains carrying CTG TNR on the leading strand mutated in one of these genes were shown to produce a different distribution of deletion products, compared to a wild type strain carrying the TNRs in the same orientation. The same was also true for a *recG* mutant strain, carrying TNR in the same orientation, but with a more pronounced effect on deletion product size. Mutation of these genes did not affect the number of deletion events taking place, but rather the size of the deletion product obtained after the event. A model was proposed in which the RecFOR, RecA and RecG proteins act on intermediates formed from instability events occurring in CAG TNR on the lagging strand template (Figure 5.16). Cells with these proteins present produce larger deletion products, while cells with mutations in the *recFOR,recA* or *recG* genes do not have such a strong bias towards large deletions.

Analysis of the effects of mutation in genes encoding helicase proteins produced some significant insights into CTG•CAG TNR instability. When the gene encoding RecQ helicase was inactivated, no significant effect was seen on deletion instability in strains carrying a TNR in either orientation relative to the origin of replication. Nor was there an effect of this mutation on expansion instability in cells carrying CTG TNR on the leading strand. However, cells carrying CAG TNR on the leading strand, displayed a severely reduced

expansion proportion in a *recQ* mutant strain – in this orientation expansions were dependent on RecQ protein.

Mutation of the gene encoding the Rep helicase resulted in an increase in deletion instability in cells carrying a CTG TNR array on the leading strand, but not in those cells with the TNR array in the opposite orientation. Deletion instability on the $(CTG)_{95}$ Δ*rep* mutant strain was similar to that observed in the wild type strain in cells in the CAG orientation, eliminating the orientation dependence of deletion instability. The CAG orientation displays a higher level of deletion instability due to the fact that the stable structure forming CTG sequence is present on the lagging strand template, which during discontinuous replication can fold into a hairpin-like structure – an intermediate in the deletion instability pathway. Strains in the CTG orientation are less unstable as the CAG sequence on the lagging strand template forms less thermodynamically stable structures, which are less likely to persist to allow deletion instability (Andreoni *et al.*, 2010; Zahra *et al.*, 2007). Mutation of the *rep* gene can lead to replication fork stalling when the oncoming fork encounter a DNA bound protein normally removed by Rep (Boubakri *et al.*, 2010). Such a stall could allow dissociation of the nascent leading strand from its template, which would allow structure formation in the CTG TNR array present on this strand in strains in the CTG orientation, leading to deletion instability. Such a dissociation event might require the action of a helicase to unwind the template and daughter strands.

Expansion instability was shown to increase in Δ*rep* mutant cells carrying CTG•CAG TNR arrays in either orientation relative to the origin of replication.

A model was proposed to explain this effect, in which replication fork arrest occurred in the *rep* mutant cells. Once arrested, the nascent leading and lagging strand are susceptible to unwinding by helicases, which providing the TNR array is present in the nascent strand could fold into secondary structure – leading to expansion instability following restart, and a further round of replication. Analysis of double mutant strains determined that in the CAG orientation RecQ was not the helicase involved in this reaction as had been hypothesized. Rather, an effect of *dinG* mutation on expansion instability was seen in *rep* mutant cells. Expansions were reduced, though not absent, in a Δ*rep*Δ*dinG* double mutant strain, suggesting that DinG helicase was required for some of the expansions seen in the *rep* mutant strain, though not all. DinG could bind the leading strand at an arrested replication fork and translocate in a 5'-3' direction unwinding the nascent and template leading strands, allowing CTG hairpins to form on the nascent strand. The observation that some expansions are still observed in a Δ*rep*Δ*dinG* double mutant, may indicate that another helicase is also involved in this expansion pathway.

The fact that expansion instability was dependent on RecQ helicase in an otherwise wild type cell, but not in a *rep* mutant strain, may be explained by different pathways occurring with different structures of arrested fork. RecQ is recruited to SSB protein. It may be that expansions in 'wild type' *E. coli* are seen in cells in which a stall has occurred leaving a gap on the leading strand template. Such a gap would be coated in SSB protein, which could then recruit RecQ to unwind the nascent leading strand. Whereas stalls arising in a *rep*

mutant may be predominantly due to the DnaB helicase being unable to proceed through duplex DNA ahead of the fork, as it remains protein bound. This could produce a stalled fork in which both leading and lagging strand synthesis stops, which would not leave an SSB coated gap on the leading strand, so not permit the intervention of RecQ.

Figure 6.1 illustrates a model in which all the described pathways of instability have been shown together. Figure 6.1a and 6.1b present the models proposed to occur in a *rep* mutant in strains carrying TNR in the CTG and CAG orientations respectively. Figure 6.1c should represent the expansion pathway in the CTG orientation in an otherwise wild type cell, while Figure 6.1e shows expansion pathway in a TNR in the CAG orientation in a wild type cell. Figure 6.1d represents the UvrD dependent deletion pathway seen in TNR in both orientations relative to the origin of replication.

The model proposes that the main factor that accounts for the difference between expansion instability in a *rep* mutant and an otherwise wild type cell, is the structure of the arrested fork. The majority of the expansion instability detected in the wild type strain is proposed to be due to blocked forks with a leading strand gap, while a *rep* mutant is proposed to produce a stall with no leading strand gap. The recruitment of specific proteins to each structure would prove to be of key importance in this model. Proteins can be recruited through a number of different mechanisms - by recognising and binding to specific DNA structures or sequences, or alternatively by other proteins either through protein-protein interactions or by being recruited to sites of post-translational

modifications. This latter situation has been shown to be the case for the UvrD homologue Srs2 in *S. cerevisiae*. The helicase recruitment is enriched by sumoylation of PCNA (Daee *et al.*, 2007; Papouli *et al.*, 2005).

Figure 6.1 also clearly shows that while models have been proposed for deletion instability in TNR arrays in both orientations, and expansion instability models have been drawn for TNR in the CAG orientation, no real model exists to explain the mechanism of expansion in the CTG orientation. Mutation of the *rep* gene was shown to increase expansion instability in cells carrying TNR arrays in this orientation. However, none of the double mutant strains tested provided evidence for helicases involved in that instability. Expansion due to helicase action in this orientation is are still predicted, however, it remains to be determined which helicases are involved. Alternatively, a flap processing pathway exists, similar to the proposed FEN1/RAD27 pathway of expansion in eukaryotes (Liu and Bambara, 2003). In this orientation the CTG sequence would be present on the nascent Okazaki fragments, making flap processing an attractive pathway for instability.

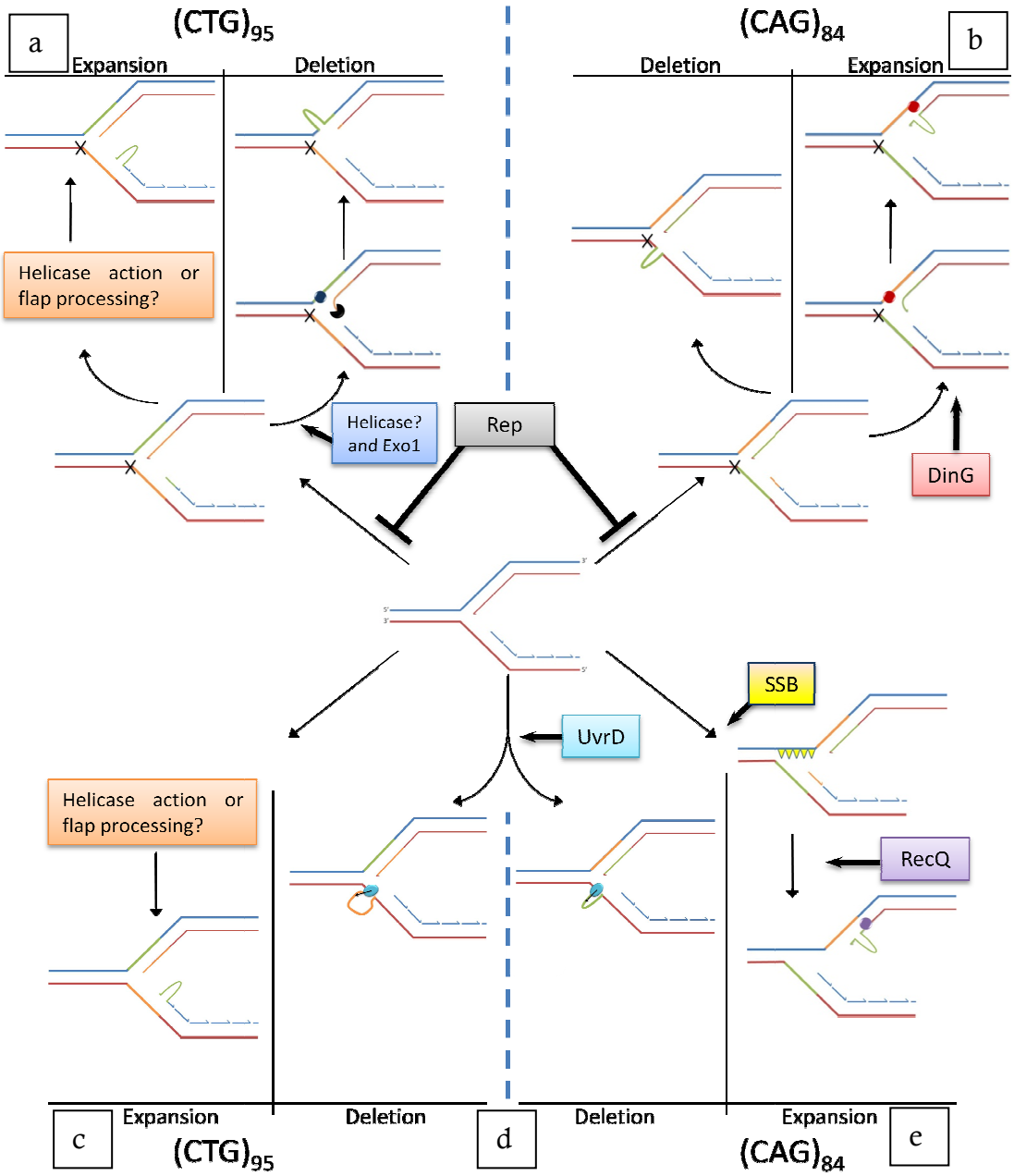

Figure 6.1: Model summary. Pathways of TNR instability in *rep* mutant strains in the CTG orientation(**a.**) and the CAG orientation (**b.**) Expansion instability pathway of CTG TNR array in an otherwise wild type cell (**c.**) UvrD dependent deletion of TNR in both orientations (**d.**) RecQ dependent expansion instability in a TNR array in the CAG orientation (**e.**)

6.2 Directions for future work

Expansion instability is of interest due to its role in TREDs in humans. Future work on this subject should focus on understanding the mechanism of expansion instability of TNR arrays. Initially testing the models I have proposed would be the most important direction for future work. In the CAG orientation of TNR arrays expansions have been proposed to arise through unwinding of the nascent leading strand. As this strand unwinds and becomes single stranded it might become a substrate for the action of the exonuclease action by Exo1/SbcB. This mechanism could explain why such small expansions were detected, as expansion instability may be affected by an equilibrium between how quickly a stable secondary structure can form in the TNR present, and the rate of nuclease action digesting the single stranded DNA. A mutant *exo1/sbcB* strain could confirm this hypothesis if this mutation increased frequency of expansions or reveals an increased average size of expansion product. The same could be true for TNR arrays in the opposite orientation. If helicase action is involved, or if a flap processing is implicated, a nuclease may be acting on the displaced nascent lagging strand before it can form into a hairpin structure. A mutation in the *recJ* gene could provide evidence of this as RecJ protein is a 5'-3' nuclease.

The proposed reason for the difference between expansion instability detected in a wild type cell and that in a *rep* mutant is the structure of the stalled fork, particularly the presence of a leading strand gap bound by SSB being able to recruit RecQ to the site in a wild type cell. An interesting early result from a

student visiting the laboratory whose project I helped to supervise, was that over-expressing SSB protein in a *rep* mutant strain increased the number of expansions detected (Els Hubertus, personal communication). If SSB protein is over-expressed it may bind more to the small gap on the leading strand at a *rep* stalled fork, or may more quickly bind the exposed template strand as helicase action unwinds the nascent strand from it, thus enabling it to recruit RecQ.

While models have been proposed for the expansion instability detected in TNR in the CAG orientation no evidence has been found to help elucidate the pathway involved in expansion instability in TNRs in the CTG orientation. This would be an important question to address, by first investigating the other helicases in the cell which could be responsible for unwinding the nascent Okazaki fragment from its template.

Using non-null alleles of helicase genes could provide useful information. Mutants in which the helicase genes in question were expressed but were deficient in their helicase function. Alternatively hyper-helicase mutants could be studied to see their effect on the instability of TNR arrays. The results in this thesis could be supported by complementing the mutations produced with a wild type gene. The work studying Δ*recQ* mutants could benefit from complementation experiments, expressing the *recQ* gene from a plasmid. Complementation might also prove interesting for the *rep* mutant strains.

More long term future work could involve investigation into the role of DinG protein. Though it has been shown *in vitro* that this protein has a substrate

specificity for forked DNA (Voloshin and Camerini-Otero, 2007) *in vivo* evidence of this could be useful in confirming it role in TNR instability. A fluorescently tagged protein could be constructed and used to see if the protein does localise at replication forks *in vivo*. Furthermore, using an approach developed in the lab to tag the *LacZ* gene it would be possible to integrate an array of *tet* operator sites, which can be bound by a fluorescently tagged Tet repressor protein, near the TNR arrays in the chromosome (White *et al.*, 2008). Using this system it could be seen whether DinG protein localises to the TNR arrays *in vivo* in *rep* mutant cells or in otherwise wild type cells. Though in this case visualization would be dependent on the number of DinG molecules involved.

Some plasmid based studies have suggested that TNRs can cause polymerase stalling *in vivo* (Krasilnikova and Mirkin, 2004; Samadashwily *et al.*, 1997) though no evidence of this on a chromosomal TNR has yet been shown. The models proposed in this thesis have been based on stochastic stalls arising in the cell rather than stalls due to the TNR arrays themselves. To address this question physical analysis by 2D gel electrophoresis could be used to look for evidence of replication stalling in the TNR arrays.

One of the main limiting issues, particularly surrounding physical analysis and microscopy of expansion instability in TNR arrays in *E. coli* is the low frequency at which the process happens. The genetic assay used here had to study expansion instability had to involve a large sample size for reliable detection of expansion instability. Whether this happens at a rate that could be

detectable or not in microscopy or other analyses could limit the effectiveness of such studies. The rate of expansion instability, and thus the potential information that could be gained from other analyses, may be increased if larger TNR arrays were used. Deletions have been shown to be length dependent (Zahra *et al.*, 2007) and it would be interesting to test if this were also true for expansion instability.

6.3 Concluding remarks

Over the last twenty years, research into the instability of TNR arrays has provided many insights into the pathways involved in this process, though much is still to be discovered. Many different model organisms have been used to study these pathways, with each providing its own benefits and limitations. *E. coli* presents a good model organism as it is a well characterized organism which can be easily manipulated genetically. However, the extent to which mechanisms discovered in the bacterium are transferable to humans is unknown.

Deletion instability predominates in *E. coli*, but human diseases are caused by expansion events. This study provides an investigation into the rare expansion events occurring in *E. coli*, and proposes a model in which helicases play a significant role in the generation of these expansions. The identity of the helicase acting varies with the structure of the replication fork acted on.

The long-term focus of TNR research has to be to understand the situation presented in human diseases and, eventually use this information to develop treatment or prevention.

References

Anderson, D.G., and Kowalczykowski, S.C. (1997). The translocating RecBCD enzyme stimulates recombination by directing RecA protein onto ssDNA in a chi-regulated manner. Cell *90*, 77-86.

Andreoni, F., Darmon, E., Poon, W.C., and Leach, D.R. (2010). Overexpression of the single-stranded DNA-binding protein (SSB) stabilises CAG*CTG triplet repeats in an orientation dependent manner. FEBS Lett *584*, 153-158.

Atkinson, J., and McGlynn, P. (2009). Replication fork reversal and the maintenance of genome stability. Nucleic Acids Res *37*, 3475-3492.

Bachrati, C.Z., and Hickson, I.D. (2003). RecQ helicases: suppressors of tumorigenesis and premature aging. Biochem J *374*, 577-606.

Baharoglu, Z., Petranovic, M., Flores, M.J., and Michel, B. (2006). RuvAB is essential for replication forks reversal in certain replication mutants. EMBO J *25*, 596-604.

Bhattacharyya, S., and Lahue, R.S. (2004). Saccharomyces cerevisiae Srs2 DNA helicase selectively blocks expansions of trinucleotide repeats. Mol Cell Biol *24*, 7324-7330.

Bhattacharyya, S., and Lahue, R.S. (2005). Srs2 helicase of Saccharomyces cerevisiae selectively unwinds triplet repeat DNA. J Biol Chem *280*, 33311-33317.

Bhattacharyya, S., Rolfsmeier, M.L., Dixon, M.J., Wagoner, K., and Lahue, R.S. (2002). Identification of RTG2 as a modifier gene for CTG*CAG repeat instability in Saccharomyces cerevisiae. Genetics *162*, 579-589.

Bichara, M., Wagner, J., and Lambert, I.B. (2006). Mechanisms of tandem repeat instability in bacteria. Mutat Res *598*, 144-163.

Bierne, H., Ehrlich, S.D., and Michel, B. (1995). Competition between parental and recombinant plasmids affects the measure of recombination frequencies. Plasmid *33*, 101-112.

Blastyak, A., Pinter, L., Unk, I., Prakash, L., Prakash, S., and Haracska, L. (2007). Yeast Rad5 protein required for postreplication repair has a DNA helicase activity specific for replication fork regression. Mol Cell *28*, 167-175.

Blattner, F.R., Plunkett, G., 3rd, Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J., Mau, B., and Shao, Y. (1997). The complete genome sequence of Escherichia coli K-12. Science *277*, 1453-1462.

Boubakri, H., de Septenville, A.L., Viguera, E., and Michel, B. (2010). The helicases DinG, Rep and UvrD cooperate to promote replication across transcription units in vivo. EMBO J *29*, 145-157.

Bowater, R.P., Rosche, W.A., Jaworski, A., Sinden, R.R., and Wells, R.D. (1996). Relationship between Escherichia coli growth and deletions of CTG.CAG triplet repeats in plasmids. J Mol Biol *264*, 82-96.

Bramhill, D., and Kornberg, A. (1988). Duplex opening by dnaA protein at novel sequences in initiation of replication at the origin of the E. coli chromosome. Cell *52*, 743-755.

Briggs, G.S., Mahdi, A.A., Weller, G.R., Wen, Q., and Lloyd, R.G. (2004). Interplay between DNA replication, recombination and repair based on the structure of RecG helicase. Philos Trans R Soc Lond B Biol Sci *359*, 49-59.

Brouwer, J.R., Willemsen, R., and Oostra, B.A. (2009). Microsatellite repeat instability and neurological disease. Bioessays *31*, 71-83.

Bugreev, D.V., Yu, X., Egelman, E.H., and Mazin, A.V. (2007). Novel pro- and anti-recombination activities of the Bloom's syndrome helicase. Genes Dev *21*, 3085-3094.

Bzymek, M., and Lovett, S.T. (2001). Evidence for two mechanisms of palindrome-stimulated deletion in Escherichia coli: single-strand annealing and replication slipped mispairing. Genetics *158*, 527-540.

Cadman, C.J., and McGlynn, P. (2004). PriA helicase and SSB interact physically and functionally. Nucleic Acids Res *32*, 6378-6387.

Campuzano, V., Montermini, L., Molto, M.D., Pianese, L., Cossee, M., Cavalcanti, F., Monros, E., Rodius, F., Duclos, F., Monticelli, A., Zara, F., Canizares, J., Koutnikova, H., Bidichandani, S.I., Gellera, C., Brice, A., Trouillas, P., De Michele, G., Filla, A., De Frutos, R., Palau, F., Patel, P.I., Di Donato, S., Mandel, J.L., Cocozza, S., Koenig, M., and Pandolfo, M. (1996). Friedreich's ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion. Science *271*, 1423-1427.

Centore, R.C., Leeson, M.C., and Sandler, S.J. (2009). UvrD303, a hyperhelicase mutant that antagonizes RecA-dependent SOS expression by a mechanism that depends on its C terminus. J Bacteriol *191*, 1429-1438.

Chong, S.S., McCall, A.E., Cota, J., Subramony, S.H., Orr, H.T., Hughes, M.R., and Zoghbi, H.Y. (1995). Gametic and somatic tissue-specific heterogeneity of the expanded SCA1 CAG repeat in spinocerebellar ataxia type 1. Nat Genet *10*, 344-350.

Chu, W.K., and Hickson, I.D. (2009). RecQ helicases: multifunctional genome caretakers. Nat Rev Cancer *9*, 644-654.

Cleary, J.D., Nichol, K., Wang, Y.H., and Pearson, C.E. (2002). Evidence of cis-acting factors in replication-mediated trinucleotide repeat instability in primate cells. Nat Genet *31*, 37-46.

Cleary, J.D., and Pearson, C.E. (2003). The contribution of cis-elements to disease-associated repeat instability: clinical and experimental evidence. Cytogenet Genome Res *100*, 25-55.

Connelly, J.C., de Leau, E.S., and Leach, D.R. (1999). DNA cleavage and degradation by the SbcCD protein complex from Escherichia coli. Nucleic Acids Res *27*, 1039-1046.

Courcelle, J., Donaldson, J.R., Chow, K.H., and Courcelle, C.T. (2003). DNA damage-induced replication fork regression and processing in Escherichia coli. Science *299*, 1064-1067.

Cromie, G.A., Connelly, J.C., and Leach, D.R. (2001). Recombination at double-strand breaks and DNA ends: conserved mechanisms from phage to humans. Mol Cell *8*, 1163-1174.

D'Hulst, C., and Kooy, R.F. (2009). Fragile X syndrome: from molecular genetics to therapy. J Med Genet *46*, 577-584.

Daee, D.L., Mertz, T., and Lahue, R.S. (2007). Postreplication repair inhibits CAG.CTG repeat expansions in Saccharomyces cerevisiae. Mol Cell Biol *27*, 102-110.

Darlow, J.M., and Leach, D.R. (1998). Secondary structures in d(CGG) and d(CCG) repeat tracts. J Mol Biol *275*, 3-16.

Dillingham, M.S., and Kowalczykowski, S.C. (2008). RecBCD enzyme and the repair of double-stranded DNA breaks. Microbiol Mol Biol Rev *72*, 642-671, Table of Contents.

Dixon, D.A., and Kowalczykowski, S.C. (1993). The recombination hotspot chi is a regulatory sequence that acts by attenuating the nuclease activity of the E. coli RecBCD enzyme. Cell *73*, 87-96.

Filippova, G.N., Thienes, C.P., Penn, B.H., Cho, D.H., Hu, Y.J., Moore, J.M., Klesert, T.R., Lobanenkov, V.V., and Tapscott, S.J. (2001). CTCF-binding sites flank CTG/CAG repeats and form a methylation-sensitive insulator at the DM1 locus. Nat Genet *28*, 335-343.

Flores, M.J., Bidnenko, V., and Michel, B. (2004). The DNA repair helicase UvrD is essential for replication fork reversal in replication mutants. EMBO Rep *5*, 983-988.

Flores, M.J., Sanchez, N., and Michel, B. (2005). A fork-clearing role for UvrD. Mol Microbiol *57*, 1664-1675.

Fojtik, P., Kejnovska, I., and Vorlickova, M. (2004). The guanine-rich fragile X chromosome repeats are reluctant to form tetraplexes. Nucleic Acids Res *32*, 298-306.

Fojtik, P., and Vorlickova, M. (2001). The fragile X chromosome (GCC) repeat folds into a DNA tetraplex at neutral pH. Nucleic Acids Res *29*, 4684-4690.

Fouche, N., Ozgur, S., Roy, D., and Griffith, J.D. (2006). Replication fork regression in repetitive DNAs. Nucleic Acids Res *34*, 6044-6050.

French, S. (1992). Consequences of replication fork movement through transcription units in vivo. Science *258*, 1362-1365.

Freudenreich, C.H., Kantrow, S.M., and Zakian, V.A. (1998). Expansion and length-dependent fragility of CTG repeats in yeast. Science *279*, 853-856.

Freudenreich, C.H., Stavenhagen, J.B., and Zakian, V.A. (1997). Stability of a CTG/CAG trinucleotide repeat in yeast is dependent on its orientation in the genome. Mol Cell Biol *17*, 2090-2098.

Gacy, A.M., Goellner, G., Juranic, N., Macura, S., and McMurray, C.T. (1995). Trinucleotide repeats that expand in human disease form hairpin structures in vitro. Cell *81*, 533-540.

Gacy, A.M., and McMurray, C.T. (1998). Influence of hairpins on template reannealing at trinucleotide repeat duplexes: a model for slipped DNA. Biochemistry *37*, 9426-9434.

Gatchel, J.R., and Zoghbi, H.Y. (2005). Diseases of unstable repeat expansion: mechanisms and common principles. Nat Rev Genet *6*, 743-755.

Gerton, J.L., and Hawley, R.S. (2005). Homologous chromosome interactions in meiosis: diversity amidst conservation. Nat Rev Genet *6*, 477-487.

Grabczyk, E., and Usdin, K. (2000). Alleviating transcript insufficiency caused by Friedreich's ataxia triplet repeats. Nucleic Acids Res *28*, 4930-4937.

Greco, C.M., Hagerman, R.J., Tassone, F., Chudley, A.E., Del Bigio, M.R., Jacquemont, S., Leehey, M., and Hagerman, P.J. (2002). Neuronal intranuclear inclusions in a new cerebellar tremor/ataxia syndrome among fragile X carriers. Brain *125*, 1760-1771.

Gregg, A.V., McGlynn, P., Jaktaji, R.P., and Lloyd, R.G. (2002). Direct rescue of stalled DNA replication forks via the combined action of PriA and RecG helicase activities. Mol Cell *9*, 241-251.

Griffin, T.J.t., and Kolodner, R.D. (1990). Purification and preliminary characterization of the Escherichia coli K-12 recF protein. J Bacteriol *172*, 6291-6299.

Gupta, M.K., Atkinson, J., and McGlynn, P. (2010). DNA structure specificity conferred on a replicative helicase by its loader. J Biol Chem *285*, 979-987.

Guy, C.P., Atkinson, J., Gupta, M.K., Mahdi, A.A., Gwynn, E.J., Rudolph, C.J., Moon, P.B., van Knippenberg, I.C., Cadman, C.J., Dillingham, M.S., Lloyd, R.G., and McGlynn, P. (2009). Rep provides a second motor at the replisome to promote duplication of protein-bound DNA. Mol Cell *36*, 654-666.

Hall, M.C., and Matson, S.W. (1999). Helicase motifs: the engine that powers DNA unwinding. Mol Microbiol *34*, 867-877.

Hanada, K., Ukita, T., Kohno, Y., Saito, K., Kato, J., and Ikeda, H. (1997). RecQ DNA helicase is a suppressor of illegitimate recombination in Escherichia coli. Proc Natl Acad Sci U S A *94*, 3860-3865.

Harmon, F.G., and Kowalczykowski, S.C. (1998). RecQ helicase, in concert with RecA and SSB proteins, initiates and disrupts DNA recombination. Genes Dev *12*, 1134-1144.

Hashem, V.I., Klysik, E.A., Rosche, W.A., and Sinden, R.R. (2002). Instability of repeated DNAs during transformation in Escherichia coli. Mutat Res *502*, 39-46.

Hashem, V.I., Rosche, W.A., and Sinden, R.R. (2004). Genetic recombination destabilizes (CTG)n.(CAG)n repeats in E. coli. Mutat Res *554*, 95-109.

Hebert, M.L., Spitz, L.A., and Wells, R.D. (2004). DNA double-strand breaks induce deletion of CTG.CAG repeats in an orientation-dependent manner in Escherichia coli. J Mol Biol *336*, 655-672.

Hebert, M.L., and Wells, R.D. (2005). Roles of double-strand breaks, nicks, and gaps in stimulating deletions of CTG.CAG repeats by intramolecular DNA repair. J Mol Biol *353*, 961-979.

Heller, R.C., and Marians, K.J. (2005a). The disposition of nascent strands at stalled replication forks dictates the pathway of replisome loading during restart. Mol Cell *17*, 733-743.

Heller, R.C., and Marians, K.J. (2005b). Unwinding of the nascent lagging strand by Rep and PriA enables the direct restart of stalled replication forks. J Biol Chem *280*, 34143-34151.

Heller, R.C., and Marians, K.J. (2006a). Replication fork reactivation downstream of a blocked nascent leading strand. Nature *439*, 557-562.

Heller, R.C., and Marians, K.J. (2006b). Replisome assembly and the direct restart of stalled replication forks. Nat Rev Mol Cell Biol *7*, 932-943.

Henricksen, L.A., Tom, S., Liu, Y., and Bambara, R.A. (2000). Inhibition of flap endonuclease 1 by flap secondary structure and relevance to repeat sequence expansion. J Biol Chem *275*, 16420-16427.

Hill, T.M., Henson, J.M., and Kuempel, P.L. (1987). The terminus region of the Escherichia coli chromosome contains two separate loci that exhibit polar inhibition of replication. Proc Natl Acad Sci U S A *84*, 1754-1758.

Hishida, T., Han, Y.W., Shibata, T., Kubota, Y., Ishino, Y., Iwasaki, H., and Shinagawa, H. (2004). Role of the Escherichia coli RecQ DNA helicase in SOS signaling and genome stabilization at stalled replication forks. Genes Dev *18*, 1886-1897.

Ho, S.N., Hunt, H.D., Horton, R.M., Pullen, J.K., and Pease, L.R. (1989). Site-directed mutagenesis by overlap extension using the polymerase chain reaction. Gene *77*, 51-59.

Hotchkiss, R.D. (1974). Models of genetic recombination. Annu Rev Microbiol *28*, 445-468.

Iyer, R.R., Pluciennik, A., Rosche, W.A., Sinden, R.R., and Wells, R.D. (2000). DNA polymerase III proofreading mutants enhance the expansion and deletion of triplet repeat sequences in Escherichia coli. J Biol Chem *275*, 2174-2184.

Jakupciak, J.P., and Wells, R.D. (2000). Gene conversion (recombination) mediates expansions of CTG[middle dot]CAG repeats. J Biol Chem *275*, 40003-40013.

Jaworski, A., Rosche, W.A., Gellibolian, R., Kang, S., Shimizu, M., Bowater, R.P., Sinden, R.R., and Wells, R.D. (1995). Mismatch repair in Escherichia coli enhances instability of (CTG)n triplet repeats from human hereditary diseases. Proc Natl Acad Sci U S A *92*, 11019-11023.

Johnson, A., and O'Donnell, M. (2005). Cellular DNA replicases: components and dynamics at the replication fork. Annu Rev Biochem *74*, 283-315.

Johzuka, K., and Ogawa, H. (1995). Interaction of Mre11 and Rad50: two proteins required for DNA repair and meiosis-specific double-strand break formation in Saccharomyces cerevisiae. Genetics *139*, 1521-1532.

Kang, S., Jaworski, A., Ohshima, K., and Wells, R.D. (1995a). Expansion and deletion of CTG repeats from human disease genes are determined by the direction of replication in E. coli. Nat Genet *10*, 213-218.

Kang, S., Ohshima, K., Shimizu, M., Amirhaeri, S., and Wells, R.D. (1995b). Pausing of DNA synthesis in vitro at specific loci in CTG and CGG triplet repeats from human hereditary disease genes. J Biol Chem *270*, 27014-27021.

Kaplan, D.L., and O'Donnell, M. (2006). RuvA is a sliding collar that protects Holliday junctions from unwinding while promoting branch migration. J Mol Biol *355*, 473-490.

Kaytor, M.D., and Orr, H.T. (2001). RNA targets of the fragile X protein. Cell *107*, 555-557.

Kerrest, A., Anand, R.P., Sundararajan, R., Bermejo, R., Liberi, G., Dujon, B., Freudenreich, C.H., and Richard, G.F. (2009). SRS2 and SGS1 prevent chromosomal breaks and stabilize triplet repeats by restraining recombination. Nat Struct Mol Biol *16*, 159-167.

Koonin, E.V. (1993). Escherichia coli dinG gene encodes a putative DNA helicase related to a group of eukaryotic helicases including Rad3 protein. Nucleic Acids Res *21*, 1497.

Kovtun, I.V., Liu, Y., Bjoras, M., Klungland, A., Wilson, S.H., and McMurray, C.T. (2007). OGG1 initiates age-dependent CAG trinucleotide expansion in somatic cells. Nature *447*, 447-452.

Kovtun, I.V., and McMurray, C.T. (2001). Trinucleotide expansion in haploid germ cells by gap repair. Nat Genet *27*, 407-411.

Kovtun, I.V., and McMurray, C.T. (2008). Features of trinucleotide repeat instability in vivo. Cell Res *18*, 198-213.

Kowalczykowski, S.C. (2000). Initiation of genetic recombination and recombination-dependent replication. Trends Biochem Sci *25*, 156-165.

Kowalczykowski, S.C., Dixon, D.A., Eggleston, A.K., Lauder, S.D., and Rehrauer, W.M. (1994). Biochemistry of homologous recombination in Escherichia coli. Microbiol Rev *58*, 401-465.

Krasilnikova, M.M., and Mirkin, S.M. (2004). Replication stalling at Friedreich's ataxia (GAA)n repeats in vivo. Mol Cell Biol *24*, 2286-2295.

Krawczun, M.S., Jenkins, E.C., and Brown, W.T. (1985). Analysis of the fragile-X chromosome: localization and detection of the fragile site in high resolution preparations. Hum Genet *69*, 209-211.

Kuzminov, A. (1999). Recombinational repair of DNA damage in Escherichia coli and bacteriophage lambda. Microbiol Mol Biol Rev *63*, 751-813, table of contents.

Lane, H.E., and Denhardt, D.T. (1975). The rep mutation. IV. Slower movement of replication forks in Escherichia coli rep strains. J Mol Biol *97*, 99-112.

Langston, L.D., and O'Donnell, M. (2006). DNA replication: keep moving and don't mind the gap. Mol Cell *23*, 155-160.

Lecointe, F., Serena, C., Velten, M., Costes, A., McGovern, S., Meile, J.C., Errington, J., Ehrlich, S.D., Noirot, P., and Polard, P. (2007). Anticipating chromosomal replication fork arrest: SSB targets repair DNA helicases to active forks. EMBO J *26*, 4239-4251.

Leeflang, E.P., Tavare, S., Marjoram, P., Neal, C.O., Srinidhi, J., MacFarlane, H., MacDonald, M.E., Gusella, J.F., de Young, M., Wexler, N.S., and Arnheim, N. (1999). Analysis of germline mutation spectra at the Huntington's disease locus supports a mitotic mutation mechanism. Hum Mol Genet *8*, 173-183.

Leeflang, E.P., Zhang, L., Tavare, S., Hubert, R., Srinidhi, J., MacDonald, M.E., Myers, R.H., de Young, M., Wexler, N.S., Gusella, J.F., and et al. (1995). Single sperm analysis of the trinucleotide repeats in the Huntington's disease gene: quantification of the mutation frequency spectrum. Hum Mol Genet *4*, 1519-1526.

Lestini, R., and Michel, B. (2008). UvrD and UvrD252 counteract RecQ, RecJ, and RecFOR in a rep mutant of Escherichia coli. J Bacteriol *190*, 5995-6001.

Lewis, L.K., Jenkins, M.E., and Mount, D.W. (1992). Isolation of DNA damage-inducible promoters in Escherichia coli: regulation of polB (dinA), dinG, and dinH by LexA repressor. J Bacteriol *174*, 3377-3385.

Li, S.H., and Li, X.J. (2004). Huntingtin-protein interactions and the pathogenesis of Huntington's disease. Trends Genet *20*, 146-154.

Libby, R.T., Hagerman, K.A., Pineda, V.V., Lau, R., Cho, D.H., Baccam, S.L., Axford, M.M., Cleary, J.D., Moore, J.M., Sopher, B.L., Tapscott, S.J., Filippova, G.N., Pearson, C.E., and La Spada, A.R. (2008). CTCF cis-regulates trinucleotide repeat instability in an epigenetic manner: a novel basis for mutational hot spot determination. PLoS Genet *4*, e1000257.

Lieber, M.R., Ma, Y., Pannicke, U., and Schwarz, K. (2003). Mechanism and regulation of human non-homologous DNA end-joining. Nat Rev Mol Cell Biol *4*, 712-720.

Link, A.J., Phillips, D., and Church, G.M. (1997). Methods for generating precise deletions and insertions in the genome of wild-type Escherichia coli: application to open reading frame characterization. J Bacteriol *179*, 6228-6237.

Liu, Y., and Bambara, R.A. (2003). Analysis of human flap endonuclease 1 mutants reveals a mechanism to prevent triplet repeat expansion. J Biol Chem *278*, 13728-13739.

Lloyd, R.G. (1991). Conjugational recombination in resolvase-deficient ruvC mutants of Escherichia coli K-12 depends on recG. J Bacteriol *173*, 5414-5418.

Lloyd, R.G., and Sharples, G.J. (1993). Dissociation of synthetic Holliday junctions by E. coli RecG protein. EMBO J *12*, 17-22.

Lopez Castel, A., Cleary, J.D., and Pearson, C.E. (2010). Repeat instability as the basis for human diseases and as a potential target for therapy. Nat Rev Mol Cell Biol *11*, 165-170.

Machwe, A., Xiao, L., Lloyd, R.G., Bolt, E., and Orren, D.K. (2007). Replication fork regression in vitro by the Werner syndrome protein (WRN): holliday junction formation, the effect of leading arm structure and a potential role for WRN exonuclease activity. Nucleic Acids Res *35*, 5729-5747.

Maisnier-Patin, S., Nordstrom, K., and Dasgupta, S. (2001). Replication arrests during a single round of replication of the Escherichia coli chromosome in the absence of DnaC activity. Mol Microbiol *42*, 1371-1382.

Manley, K., Shirley, T.L., Flaherty, L., and Messer, A. (1999). Msh2 deficiency prevents in vivo somatic instability of the CAG repeat in Huntington disease transgenic mice. Nat Genet *23*, 471-473.

Marszalek, J., and Kaguni, J.M. (1994). DnaA protein directs the binding of DnaB protein in initiation of DNA replication in Escherichia coli. J Biol Chem *269*, 4883-4890.

Matson, S.W., and Kaiser-Rogers, K.A. (1990). DNA helicases. Annu Rev Biochem *59*, 289-329.

McGlynn, P., and Lloyd, R.G. (2001). Rescue of stalled replication forks by RecG: simultaneous translocation on the leading and lagging strand templates supports an active DNA unwinding model of fork reversal and Holliday junction formation. Proc Natl Acad Sci U S A *98*, 8227-8234.

McInerney, P., Johnson, A., Katz, F., and O'Donnell, M. (2007). Characterization of a triple DNA polymerase replisome. Mol Cell *27*, 527-538.

Merlin, C., McAteer, S., and Masters, M. (2002). Tools for characterization of Escherichia coli genes of unknown function. J Bacteriol *184*, 4573-4581.

Michel, B. (2000). Replication fork arrest and DNA recombination. Trends Biochem Sci *25*, 173-178.

Michel, B., Grompone, G., Flores, M.J., and Bidnenko, V. (2004). Multiple pathways process stalled replication forks. Proc Natl Acad Sci U S A *101*, 12783-12788.

Miret, J.J., Pessoa-Brandao, L., and Lahue, R.S. (1997). Instability of CAG and CTG trinucleotide repeats in Saccharomyces cerevisiae. Mol Cell Biol *17*, 3382-3387.

Miret, J.J., Pessoa-Brandao, L., and Lahue, R.S. (1998). Orientation-dependent and sequence-specific expansions of CTG/CAG trinucleotide repeats in Saccharomyces cerevisiae. Proc Natl Acad Sci U S A *95*, 12438-12443.

Mirkin, S.M. (2007). Expandable DNA repeats and human disease. Nature *447*, 932-940.

Mitas, M. (1997). Trinucleotide repeats associated with human disease. Nucleic Acids Res *25*, 2245-2254.

Mitas, M., Yu, A., Dill, J., and Haworth, I.S. (1995a). The trinucleotide repeat sequence d(CGG)15 forms a heat-stable hairpin containing Gsyn. Ganti base pairs. Biochemistry *34*, 12803-12811.

Mitas, M., Yu, A., Dill, J., Kamp, T.J., Chambers, E.J., and Haworth, I.S. (1995b). Hairpin properties of single-stranded DNA containing a GC-rich triplet repeat: (CTG)15. Nucleic Acids Res *23*, 1050-1059.

Morag, A.S., Saveson, C.J., and Lovett, S.T. (1999). Expansion of DNA repeats in Escherichia coli: effects of recombination and replication functions. J Mol Biol *289*, 21-27.

Morimatsu, K., and Kowalczykowski, S.C. (2003). RecFOR proteins load RecA protein onto gapped DNA to accelerate DNA strand exchange: a universal step of recombinational repair. Mol Cell *11*, 1337-1347.

Mott, M.L., and Berger, J.M. (2007). DNA replication initiation: mechanisms and regulation in bacteria. Nat Rev Microbiol *5*, 343-354.

Mulcair, M.D., Schaeffer, P.M., Oakley, A.J., Cross, H.F., Neylon, C., Hill, T.M., and Dixon, N.E. (2006). A molecular mousetrap determines polarity of termination of DNA replication in E. coli. Cell *125*, 1309-1319.

Mutsuddi, M., and Rebay, I. (2005). Molecular genetics of spinocerebellar ataxia type 8 (SCA8). RNA Biol *2*, 49-52.

Nakayama, K., Irino, N., and Nakayama, H. (1985). The recQ gene of Escherichia coli K12: molecular cloning and isolation of insertion mutants. Mol Gen Genet *200*, 266-271.

Napierala, M., and Krzyzosiak, W.J. (1997). CUG repeats present in myotonin kinase RNA form metastable "slippery" hairpins. J Biol Chem *272*, 31079-31085.

Napierala, M., Parniewski, P., Pluciennik, A., and Wells, R.D. (2002). Long CTG.CAG repeat sequences markedly stimulate intramolecular recombination. J Biol Chem *277*, 34087-34100.

Nenguke, T., Aladjem, M.I., Gusella, J.F., Wexler, N.S., and Arnheim, N. (2003). Candidate DNA replication initiation regions at human trinucleotide repeat disease loci. Hum Mol Genet *12*, 1021-1028.

Orr, H.T., and Zoghbi, H.Y. (2007). Trinucleotide repeat disorders. Annu Rev Neurosci *30*, 575-621.

Pages, V., and Fuchs, R.P. (2003). Uncoupling of leading- and lagging-strand DNA replication during lesion bypass in vivo. Science *300*, 1300-1303.

Papouli, E., Chen, S., Davies, A.A., Huttner, D., Krejci, L., Sung, P., and Ulrich, H.D. (2005). Crosstalk between SUMO and ubiquitin on PCNA is mediated by recruitment of the helicase Srs2p. Mol Cell *19*, 123-133.

Pearson, C.E., Nichol Edamura, K., and Cleary, J.D. (2005). Repeat instability: mechanisms of dynamic mutations. Nat Rev Genet *6*, 729-742.

Pearson, C.E., and Sinden, R.R. (1998). Trinucleotide repeat DNA structures: dynamic mutations from dynamic DNA. Curr Opin Struct Biol *8*, 321-330.

Pearson, C.E., Tam, M., Wang, Y.H., Montgomery, S.E., Dar, A.C., Cleary, J.D., and Nichol, K. (2002). Slipped-strand DNAs formed by long (CAG)*(CTG) repeats: slipped-out repeats and slip-out junctions. Nucleic Acids Res *30*, 4534-4547.

Pelletier, R., Krasilnikova, M.M., Samadashwily, G.M., Lahue, R., and Mirkin, S.M. (2003). Replication and expansion of trinucleotide repeats in yeast. Mol Cell Biol *23*, 1349-1357.

Petit, M.A., and Ehrlich, D. (2002). Essential bacterial helicases that counteract the toxicity of recombination proteins. EMBO J *21*, 3137-3147.

Petruska, J., Arnheim, N., and Goodman, M.F. (1996). Stability of intrastrand hairpin structures formed by the CAG/CTG class of DNA triplet repeats associated with neurological diseases. Nucleic Acids Res *24*, 1992-1998.

Pluciennik, A., Iyer, R.R., Napierala, M., Larson, J.E., Filutowicz, M., and Wells, R.D. (2002). Long CTG.CAG repeats from myotonic dystrophy are preferred sites for intermolecular recombination. J Biol Chem *277*, 34074-34086.

Possoz, C., Filipe, S.R., Grainge, I., and Sherratt, D.J. (2006). Tracking of controlled Escherichia coli replication fork stalling and restart at repressor-bound DNA in vivo. EMBO J *25*, 2596-2604.

Postow, L., Ullsperger, C., Keller, R.W., Bustamante, C., Vologodskii, A.V., and Cozzarelli, N.R. (2001). Positive torsional strain causes the formation of a four-way junction at replication forks. J Biol Chem *276*, 2790-2796.

Potaman, V.N., Oussatcheva, E.A., Lyubchenko, Y.L., Shlyakhtenko, L.S., Bidichandani, S.I., Ashizawa, T., and Sinden, R.R. (2004). Length-dependent structure formation in Friedreich ataxia (GAA)n*(TTC)n repeats at neutral pH. Nucleic Acids Res *32*, 1224-1231.

Reyes-Lamothe, R., Sherratt, D.J., and Leake, M.C. Stoichiometry and architecture of active DNA replication machinery in Escherichia coli. Science *328*, 498-501.

Richards, R.I. (2001). Dynamic mutations: a decade of unstable expanded repeats in human genetic disease. Hum Mol Genet *10*, 2187-2194.

Richards, R.I., and Sutherland, G.R. (1992). Dynamic mutations: a new class of mutations causing human disease. Cell *70*, 709-712.

Rolfsmeier, M.L., Dixon, M.J., Pessoa-Brandao, L., Pelletier, R., Miret, J.J., and Lahue, R.S. (2001). Cis-elements governing trinucleotide repeat instability in Saccharomyces cerevisiae. Genetics *157*, 1569-1579.

Rolfsmeier, M.L., and Lahue, R.S. (2000). Stabilizing effects of interruptions on trinucleotide repeat expansions in Saccharomyces cerevisiae. Mol Cell Biol *20*, 173-180.

Rudolph, C.J., Dhillon, P., Moore, T., and Lloyd, R.G. (2007). Avoiding and resolving conflicts between DNA replication and transcription. DNA Repair (Amst) *6*, 981-993.

Samadashwily, G.M., Raca, G., and Mirkin, S.M. (1997). Trinucleotide repeats affect DNA replication in vivo. Nat Genet *17*, 298-304.

Sandler, S.J. (2000). Multiple genetic pathways for restarting DNA replication forks in Escherichia coli K-12. Genetics *155*, 487-497.

Sandler, S.J., and Marians, K.J. (2000). Role of PriA in replication fork reactivation in Escherichia coli. J Bacteriol *182*, 9-13.

Savouret, C., Brisson, E., Essers, J., Kanaar, R., Pastink, A., te Riele, H., Junien, C., and Gourdon, G. (2003). CTG repeat instability and size variation timing in DNA repair-deficient mice. EMBO J *22*, 2264-2273.

Schofield, M.J., and Hsieh, P. (2003). DNA mismatch repair: molecular mechanisms and biological function. Annu Rev Microbiol *57*, 579-608.

Schumacher, S., Fuchs, R.P., and Bichara, M. (1998). Expansion of CTG repeats from human disease genes is dependent upon replication mechanisms in Escherichia coli: the effect of long patch mismatch repair revisited. J Mol Biol *279*, 1101-1110.

Schweitzer, J.K., and Livingston, D.M. (1997). Destabilization of CAG trinucleotide repeat tracts by mismatch repair mutations in yeast. Hum Mol Genet *6*, 349-355.

Schweitzer, J.K., and Livingston, D.M. (1998). Expansions of CAG repeat tracts are frequent in a yeast mutant defective in Okazaki fragment maturation. Hum Mol Genet *7*, 69-74.

Seigneur, M., Bidnenko, V., Ehrlich, S.D., and Michel, B. (1998). RuvAB acts at arrested replication forks. Cell *95*, 419-430.

Seigneur, M., Ehrlich, S.D., and Michel, B. (2000). RuvABC-dependent double-strand breaks in dnaBts mutants require recA. Mol Microbiol *38*, 565-574.

Sharples, G.J., and Leach, D.R. (1995). Structural and functional similarities between the SbcCD proteins of Escherichia coli and the RAD50 and MRE11 (RAD32) recombination and repair proteins of yeast. Mol Microbiol *17*, 1215-1217.

Shereda, R.D., Bernstein, D.A., and Keck, J.L. (2007). A central role for SSB in Escherichia coli RecQ DNA helicase function. J Biol Chem *282*, 19247-19258.

Sinden, R.R., Potaman, V.N., Oussatcheva, E.A., Pearson, C.E., Lyubchenko, Y.L., and Shlyakhtenko, L.S. (2002). Triplet repeat DNA structures and human genetic disease: dynamic mutations from dynamic DNA. J Biosci *27*, 53-65.

Singleton, M.R., Dillingham, M.S., and Wigley, D.B. (2007). Structure and mechanism of helicases and nucleic acid translocases. Annu Rev Biochem *76*, 23-50.

Sopher, B.L., Myrick, S.B., Hong, J.Y., Smith, A.C., and La Spada, A.R. (2000). In vivo expansion of trinucleotide repeats yields plasmid and YAC constructs for targeting and transgenesis. Gene *261*, 383-390.

Spiro, C., Pelletier, R., Rolfsmeier, M.L., Dixon, M.J., Lahue, R.S., Gupta, G., Park, M.S., Chen, X., Mariappan, S.V., and McMurray, C.T. (1999). Inhibition of FEN-1 processing by DNA secondary structure at trinucleotide repeats. Mol Cell *4*, 1079-1085.

Stukenberg, P.T., Turner, J., and O'Donnell, M. (1994). An explanation for lagging strand replication: polymerase hopping among DNA sliding clamps. Cell *78*, 877-887.

Sundararajan, R., Gellon, L., Zunder, R.M., and Freudenreich, C.H. (2010). Double-strand break repair pathways protect against CAG/CTG repeat expansions, contractions and repeat-mediated chromosomal fragility in Saccharomyces cerevisiae. Genetics *184*, 65-77.

Taylor, A.F., and Smith, G.R. (1985). Substrate specificity of the DNA unwinding activity of the RecBC enzyme of Escherichia coli. J Mol Biol *185*, 431-443.

Tian, L., Hou, C., Tian, K., Holcomb, N.C., Gu, L., and Li, G.M. (2009). Mismatch recognition protein MutSbeta does not hijack (CAG)n hairpin repair in vitro. J Biol Chem *284*, 20452-20456.

Tuteja, N., and Tuteja, R. (2004). Prokaryotic and eukaryotic DNA helicases. Essential molecular motor proteins for cellular machinery. Eur J Biochem *271*, 1835-1848.

Van Esch, H. (2006). The Fragile X premutation: new insights and clinical consequences. Eur J Med Genet *49*, 1-8.

Veaute, X., Jeusset, J., Soustelle, C., Kowalczykowski, S.C., Le Cam, E., and Fabre, F. (2003). The Srs2 helicase prevents recombination by disrupting Rad51 nucleoprotein filaments. Nature *423*, 309-312.

Verkerk, A.J., Pieretti, M., Sutcliffe, J.S., Fu, Y.H., Kuhl, D.P., Pizzuti, A., Reiner, O., Richards, S., Victoria, M.F., Zhang, F.P., and et al. (1991). Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. Cell *65*, 905-914.

Voloshin, O.N., and Camerini-Otero, R.D. (2007). The DinG protein from Escherichia coli is a structure-specific helicase. J Biol Chem *282*, 18437-18447.

Voloshin, O.N., Vanevski, F., Khil, P.P., and Camerini-Otero, R.D. (2003). Characterization of the DNA damage-inducible helicase DinG from Escherichia coli. J Biol Chem *278*, 28284-28293.

Webb, B.L., Cox, M.M., and Inman, R.B. (1997). Recombinational DNA repair: the RecF and RecR proteins limit the extension of RecA filaments beyond single-strand DNA gaps. Cell *91*, 347-356.

Whitby, M.C., Vincent, S.D., and Lloyd, R.G. (1994). Branch migration of Holliday junctions: identification of RecG protein as a junction specific DNA helicase. EMBO J *13*, 5220-5228.

White, M.A., Eykelenboom, J.K., Lopez-Vernaza, M.A., Wilson, E., and Leach, D.R. (2008). Non-random segregation of sister chromosomes in Escherichia coli. Nature *455*, 1248-1250.

Yancey-Wrona, J.E., and Matson, S.W. (1992). Bound Lac repressor protein differentially inhibits the unwinding reactions catalyzed by DNA helicases. Nucleic Acids Res *20*, 6713-6721.

Yao, N.Y., and O'Donnell, M. (2009). Replisome structure and conformational dynamics underlie fork progression past obstacles. Curr Opin Cell Biol *21*, 336-343.

Yoon, S.R., Dubeau, L., de Young, M., Wexler, N.S., and Arnheim, N. (2003). Huntington disease expansion mutations in humans can occur before meiosis is completed. Proc Natl Acad Sci U S A *100*, 8834-8838.

Yu, A., Dill, J., and Mitas, M. (1995). The purine-rich trinucleotide repeat sequences d(CAG)15 and d(GAC)15 form hairpins. Nucleic Acids Res *23*, 4055-4057.

Zahra, R., Blackwood, J.K., Sales, J., and Leach, D.R. (2007). Proofreading and secondary structure processing determine the orientation dependence of CAG x CTG trinucleotide repeat instability in Escherichia coli. Genetics *176*, 27-41.

Zhang, X.D., Dou, S.X., Xie, P., Hu, J.S., Wang, P.Y., and Xi, X.G. (2006). Escherichia coli RecQ is a rapid, efficient, and monomeric helicase. J Biol Chem *281*, 12655-12663.

Zieg, J., Maples, V.F., and Kushner, S.R. (1978). Recombinant levels of Escherichia coli K-12 mutants deficient in various replication, recombination, or repair genes. J Bacteriol *134*, 958-966.

Zoghbi, H.Y., and Orr, H.T. (2000). Glutamine repeats and neurodegeneration. Annu Rev Neurosci *23*, 217-247.